Statistics for Astronomers: Lecture 4, 2020.09.30

Prof. Sundar Srinivasan

IRyA/UNAM

Review

Variance and covariance.
Uncorrelated random numbers & Bienaymé's Identity.
Distributions: Bernoulli & Binomial.

Homework #2: erratum

In Problem 5, the velocities projected along the line of sight should be $v_{\rm rad} \cos \phi$, not $v_{\rm rad} \sin \phi$.

Clarification

Three quantities:

Population mean μ – constant.

$\mathbb{E}[X]$ – constant, estimator for μ.

Sample mean $\bar{X}$ – random variable, also estimator for μ.

Law of Large Numbers: as $N \to \infty$, $\bar{X} \to \mathbb{E}[X]$.

Central Limit Theorem: as $N \to \infty$, $\bar{X} \to \mu$, if X_i mutually independent.

In general, the three numbers are not equal. We have three deviations to investigate:

1. $\mathbb{E}[X] - \mu$, the bias, constant for a distribution.
2. $X - \mathbb{E}[X]$, a random variable. Convert to constant by computing $\mathbb{E}[(X - \mathbb{E}[X])^2]$, the variance.
3. $X - \mu$, another random variable. Convert to constant by computing $\mathbb{E}[(X - \mu)^2]$, the mean square error (MSE).

We'll come back to this when discussing the Bias-Variance Tradeoff.

Poisson (discrete, `scipy.stats.poisson`)

Probability distribution of the number of events in a fixed interval, such that the events

1. are rare (interval of observation $\times$ occurrence rate $\ll 1$),
2. are mutually independent, and
3. occur at a constant average rate independent of the location of the interval.

Poisson (discrete, `scipy.stats.poisson`)

Probability distribution of the number of events in a fixed interval, such that the events

1. are rare (interval of observation $\times$ occurrence rate $\ll 1$),
2. are mutually independent, and
3. occur at a constant average rate independent of the location of the interval.

Poisson = Binomial with $p \ll 1$ and $n \gg k$ (# trials $\gg$ # successes), such that np is finite.
$\lambda \equiv np$ is the mean number of events per interval.

Poisson (discrete, `scipy.stats.poisson`)

Probability distribution of the number of events in a fixed interval, such that the events

1. are rare (interval of observation $\times$ occurrence rate $\ll 1$),
2. are mutually independent, and
3. occur at a constant average rate independent of the location of the interval.

Poisson $=$ Binomial with $p \ll 1$ and $n \gg k$ (# trials $\gg$ # successes), such that np is finite.
$\lambda \equiv np$ is the mean number of events per interval.

Examples: the probability that

two supernovae go off in the Milky Way within the next 100 years,
a sample of ^{137}Cs nuclei produces 15 decays in the next minute *rare?*.
a mag 7.0 earthquake hits Mexico City within the next ten years *[independent?]*,
3 photons from a target will hit a telescope detector within the next second,

One of the earliest applications: the ▸ Prussian army's "death by horse kick" data

Poisson (discrete, `scipy.stats.poisson`)

Probability distribution of the number of events in a fixed interval, such that the events

1. are rare (interval of observation $\times$ occurrence rate $\ll 1$),
2. are mutually independent, and
3. occur at a constant average rate independent of the location of the interval.

Poisson = Binomial with $p \ll 1$ and $n \gg k$ (# trials $\gg$ # successes), such that np is finite.
$\lambda \equiv np$ is the mean number of events per interval.

Examples: the probability that

two supernovae go off in the Milky Way within the next 100 years,
a sample of ^{137}Cs nuclei produces 15 decays in the next minute *rare?*.
a mag 7.0 earthquake hits Mexico City within the next ten years *[independent?]*,
3 photons from a target will hit a telescope detector within the next second,

One of the earliest applications: the ▸ Prussian army's "death by horse kick" data

$$\mathrm{Poisson}(k, \lambda) = \lambda^k \frac{e^{-\lambda}}{k!}$$ (Can be derived from Binomial PMF by setting $\lambda \equiv np$ then $n \to \infty$).

Poisson (discrete, `scipy.stats.poisson`)

Probability distribution of the number of events in a fixed interval, such that the events

1. are rare (interval of observation $\times$ occurrence rate $\ll 1$),
2. are mutually independent, and
3. occur at a constant average rate independent of the location of the interval.

Poisson = Binomial with $p \ll 1$ and $n \gg k$ (# trials $\gg$ # successes), such that np is finite.
$\lambda \equiv np$ is the mean number of events per interval.

Examples: the probability that

two supernovae go off in the Milky Way within the next 100 years,
a sample of ^{137}Cs nuclei produces 15 decays in the next minute *rare?*.
a mag 7.0 earthquake hits Mexico City within the next ten years *[independent?]*,
3 photons from a target will hit a telescope detector within the next second,

One of the earliest applications: the ▸ Prussian army's "death by horse kick" data

$$\text{Poisson}(k, \lambda) = \lambda^k \frac{e^{-\lambda}}{k!}$$ (Can be derived from Binomial PMF by setting $\lambda \equiv np$ then $n \to \infty$).

Number of independent, rare events in a fixed interval: Poisson distribution.
interval between consecutive Poisson events: Exponential distribution.

Poisson distribution, contd.

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\lambda^k \frac{e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda} e^{\lambda} = \lambda (= np)$$

Poisson distribution, contd.

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\lambda^k \frac{e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda}\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\right) = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda} e^{\lambda} = \lambda (= np)$$

Similarly, $Var(X) = \lambda$ Compare with $\mathrm{Binomial}(n, p)$: $Var(X) = np(1-p)$, with $p \ll 1$.

$\Longrightarrow$ Measurement of N Poisson events: standard deviation (an "uncertainty") of $\sqrt{N}$.

Poisson distribution, contd.

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\lambda^k \frac{e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda}\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\right) = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda} e^{\lambda} = \lambda (= np)$$

Similarly, $Var(X) = \lambda$ Compare with $\mathrm{Binomial}(n, p)$: $Var(X) = np(1-p)$, with $p \ll 1$.

$\Longrightarrow$ Measurement of N Poisson events: standard deviation (an "uncertainty") of $\sqrt{N}$.

Illustration: A telescope receives an average of S photons s^{-1} from a source of interest. The average background contribution is B photons s^{-1}. What signal-to-noise ratio is obtained from a 10-second observation?

Poisson distribution, contd.

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\lambda^k \frac{e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda}\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\right) = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda} e^{\lambda} = \lambda (= np)$$

Similarly, $Var(X) = \lambda$ Compare with $\mathrm{Binomial}(n, p)$: $Var(X) = np(1-p)$, with $p \ll 1$.

$\Longrightarrow$ Measurement of N Poisson events: standard deviation (an "uncertainty") of $\sqrt{N}$.

Illustration: A telescope receives an average of S photons s^{-1} from a source of interest. The average background contribution is B photons s^{-1}. What signal-to-noise ratio is obtained from a 10-second observation?

Source and background photons satisfy conditions for Poisson statistics –
in both cases, the emission of each photon is independent of the emission of other photons.

Poisson distribution, contd.

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\lambda^k \frac{e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} = \lambda e^{-\lambda} \frac{\partial}{\partial \lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) = \lambda e^{-\lambda} \frac{\partial}{\partial \lambda} e^{\lambda} = \lambda (= np)$$

Similarly, $Var(X) = \lambda$ Compare with $\mathrm{Binomial}(n, p)$: $Var(X) = np(1-p)$, with $p \ll 1$.

$\Longrightarrow$ Measurement of N Poisson events: standard deviation (an "uncertainty") of $\sqrt{N}$.

Illustration: A telescope receives an average of S photons s^{-1} from a source of interest. The average background contribution is B photons s^{-1}. What signal-to-noise ratio is obtained from a 10-second observation?

Source and background photons satisfy conditions for Poisson statistics – in both cases, the emission of each photon is independent of the emission of other photons.

Expected value of the source counts: $\lambda_S = 10 \times S$; background counts: $\lambda_B = 10 \times B$.

The corresponding variances: $10 \times S$ and $10 \times B$.

Poisson distribution, contd.

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\lambda^k \frac{e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \right) = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda} e^{\lambda} = \lambda (= np)$$

Similarly, $Var(X) = \lambda$ Compare with $\mathrm{Binomial}(n, p)$: $Var(X) = np(1-p)$, with $p \ll 1$.

$\Longrightarrow$ Measurement of N Poisson events: standard deviation (an "uncertainty") of $\sqrt{N}$.

Illustration: A telescope receives an average of S photons s^{-1} from a source of interest. The average background contribution is B photons s^{-1}. What signal-to-noise ratio is obtained from a 10-second observation?

Source and background photons satisfy conditions for Poisson statistics – in both cases, the emission of each photon is independent of the emission of other photons.

Expected value of the source counts: $\lambda_S = 10 \times S$; background counts: $\lambda_B = 10 \times B$.

The corresponding variances: $10 \times S$ and $10 \times B$.

Variance on total photons observed in 10 s:

$10 \times (S + B)$ (source and background are uncorrelated).

Poisson distribution, contd.

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\lambda^k \frac{e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda}\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\right) = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda} e^{\lambda} = \lambda (= np)$$

Similarly, $Var(X) = \lambda$ Compare with $\mathrm{Binomial}(n, p)$: $Var(X) = np(1-p)$, with $p \ll 1$.

$\Longrightarrow$ Measurement of N Poisson events: standard deviation (an "uncertainty") of $\sqrt{N}$.

Illustration: A telescope receives an average of S photons s^{-1} from a source of interest. The average background contribution is B photons s^{-1}. What signal-to-noise ratio is obtained from a 10-second observation?

Source and background photons satisfy conditions for Poisson statistics – in both cases, the emission of each photon is independent of the emission of other photons.

Expected value of the source counts: $\lambda_S = 10 \times S$; background counts: $\lambda_B = 10 \times B$.

The corresponding variances: $10 \times S$ and $10 \times B$.

Variance on total photons observed in 10 s:

$10 \times (S + B)$ (source and background are uncorrelated).

$\Longrightarrow$ signal-to-noise of a 10-s observation is $\sqrt{10}\dfrac{S}{\sqrt{S+B}}$.

Poisson distribution, contd.

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\lambda^k \frac{e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{k\lambda^{k-1}}{k!} = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda}\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\right) = \lambda e^{-\lambda} \frac{\partial}{\partial\lambda} e^{\lambda} = \lambda (= np)$$

Similarly, $Var(X) = \lambda$ Compare with $\mathrm{Binomial}(n, p)$: $Var(X) = np(1-p)$, with $p \ll 1$.

$\implies$ Measurement of N Poisson events: standard deviation (an "uncertainty") of $\sqrt{N}$.

Illustration: A telescope receives an average of S photons s^{-1} from a source of interest. The average background contribution is B photons s^{-1}. What signal-to-noise ratio is obtained from a 10-second observation?

Source and background photons satisfy conditions for Poisson statistics – in both cases, the emission of each photon is independent of the emission of other photons.

Expected value of the source counts: $\lambda_S = 10 \times S$; background counts: $\lambda_B = 10 \times B$.

The corresponding variances: $10 \times S$ and $10 \times B$.

Variance on total photons observed in 10 s:

$10 \times (S + B)$ (source and background are uncorrelated).

$\implies$ signal-to-noise of a 10-s observation is $\sqrt{10}\dfrac{S}{\sqrt{S+B}}$.

Due to Poisson statistics, increasing the exposure time by a factor f increases the S/N by $\sqrt{f}$.

Uniform (continuous, `scipy.stats.uniform`)

Const. prob. per unit interval $\Rightarrow p_X(x) = \dfrac{1}{b-a}\mathbb{I}_{\{a \leq x \leq b\}}(x)$ (normalisation requires finite interval!)

Uniform (continuous, `scipy.stats.uniform`)

Const. prob. per unit interval $\Rightarrow p_X(x) = \frac{1}{b-a}\mathbb{I}_{\{a \leq x \leq b\}}(x)$ (normalisation requires finite interval!)

$X \sim \mathrm{Uniform}(a, b); X = a + (b-a)Y$, where $Y \sim \mathrm{Uniform}(0, 1)$ standard uniform distribution.

Uniform (continuous, `scipy.stats.uniform`)

Const. prob. per unit interval $\Rightarrow p_X(x) = \frac{1}{b-a}\mathbb{I}_{\{a\le x\le b\}}(x)$ (normalisation requires finite interval!)

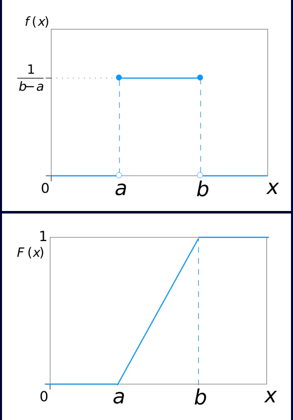
$X \sim \mathrm{Uniform}(a,b); X = a + (b-a)Y$, where $Y \sim \mathrm{Uniform}(0,1)$ standard uniform distribution.

How do you generate $\mathrm{Uniform}(4,10)$? Use the `loc` and `scale` keywords (see documentation).

Uniform (continuous, `scipy.stats.uniform`)

Const. prob. per unit interval $\Rightarrow p_X(x) = \frac{1}{b-a}\mathbb{I}_{\{a \le x \le b\}}(x)$ (normalisation requires finite interval!)

$X \sim \mathrm{Uniform}(a, b); X = a + (b-a)Y$, where $Y \sim \mathrm{Uniform}(0, 1)$ standard uniform distribution.



PDF (top) and CDF (bottom). Credit:user:IkamusumeFan/CC BY-SA 3.0

How do you generate $\mathrm{Uniform}(4, 10)$? Use the `loc` and `scale` keywords (see documentation).

$$\mathbb{E}[X] = \frac{a+b}{2} = \text{median} = \text{mode (symmetric distribution)}.$$

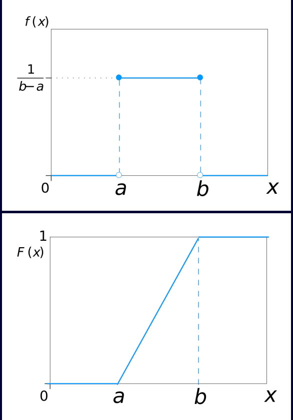
$$\mathrm{Var}(X) = \frac{1}{12}(a+b)^2$$

CDF: $F_X(x) = \frac{x-a}{b-a}$

Uniform (continuous, `scipy.stats.uniform`)

Const. prob. per unit interval $\Rightarrow p_X(x) = \frac{1}{b-a}\mathbb{I}_{\{a \le x \le b\}}(x)$ (normalisation requires finite interval!)

$X \sim \mathrm{Uniform}(a, b); X = a + (b-a)Y$, where $Y \sim \mathrm{Uniform}(0, 1)$ standard uniform distribution.



PDF (top) and CDF (bottom).
Credit:user:IkamusumeFan/CC BY-SA 3.0

How do you generate $\mathrm{Uniform}(4, 10)$? Use the `loc` and `scale` keywords (see documentation).

$$\mathbb{E}[X] = \frac{a+b}{2} = \text{median} = \text{mode (symmetric distribution).}$$

$$\mathrm{Var}(X) = \frac{1}{12}(a+b)^2$$

$$\text{CDF: } F_X(x) = \frac{x-a}{b-a}$$

Multiple samples from `scipy.stats.uniform` are uncorrelated:

```
from scipy.stats import uniform
x = uniform.rvs(size = [2, 1000]) # 2 random vectors
print(np.corrcoef(x)[0, 1])        # should be small
```

Exponential (continuous, `scipy.stats.expon`)

Number of independent, rare events in a fixed interval: Poisson distribution.
Interval between consecutive Poisson events: Exponential distribution.

Examples:
- Time between consecutive decays of a radionuclide.
- Time until next bus arrives.

PDF: $p_X(x) \equiv \mathrm{Exponential}[\lambda] = \lambda e^{-\lambda x}$.

Waiting Time Paradox see here.
- The Exponential distribution is memoryless: $P(T > t + s | P > s) = P(T > t)$.
- No memory of already having waited for time s.
- At any time, probability that next event will occur after time $t = \lambda e^{-\lambda t}$.

Mean interval between successive Poisson events = inverse of the mean rate of occurrence:

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

Variance: $\mathrm{Var}(X) = \frac{1}{\lambda^2}$

Normal (continuous, `scipy.stats.normal`)

Standard normal distribution

PDF: $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$ (mean: 0, variance: 1).

Normal (continuous, `scipy.stats.normal`)

Standard normal distribution

PDF: $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$ (mean: 0, variance: 1).

CDF : $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{t^2}{2}\right] dt$

Normal (continuous, `scipy.stats.normal`)

Standard normal distribution

PDF: $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$ (mean: 0, variance: 1).

$$\text{CDF}: \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{t^2}{2}\right] dt \quad = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right],$$

where $\text{erf}(x)$ is the error function: $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} \exp\left[-t^2\right] dt$.

Normal (continuous, `scipy.stats.normal`)

Standard normal distribution

PDF: $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$ (mean: 0, variance: 1).

CDF : $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{t^2}{2}\right] dt \quad = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$,

where $\text{erf}(x)$ is the error function: $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} \exp\left[-t^2\right] dt$.

Shifted $(0 \to \mu)$ and scaled $(1 \to \sigma^2)$ distribution:

Set $x = \mu z + \sigma$ (`loc` and `scale` in `stats.normal`) to get the general PDF:

$$\mathcal{N}(\mu, \sigma^2) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \Longrightarrow \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sigma}\, \varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}\, \varphi(z).$$

$Z \sim \varphi(z)$ (standard normal deviate) $\Longleftrightarrow$ $X \equiv \sigma Z + \mu \sim \mathcal{N}(\mu, \sigma)$ (normal deviate).

Normal (continuous, `scipy.stats.normal`)

Standard normal distribution

PDF: $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$ (mean: 0, variance: 1).

$$\text{CDF}: \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{t^2}{2}\right] dt \quad = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right],$$

where $\text{erf}(x)$ is the error function: $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\left[-t^2\right] dt$.

Shifted ($0 \to \mu$) and scaled ($1 \to \sigma^2$) distribution:

Set $x = \mu z + \sigma$ (`loc` and `scale` in `stats.normal`) to get the general PDF:

$$\mathcal{N}(\mu, \sigma^2) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \implies \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}\varphi(z).$$

$Z \sim \varphi(z)$ (standard normal deviate) $\iff X \equiv \sigma Z + \mu \sim \mathcal{N}(\mu, \sigma)$ (normal deviate).

Standardisation: the procedure of transforming a random variable into a location- and scale-free version of itself (subtract the mean, divide by the standard deviation).

(A lot) more on the normal distribution later...

Central Limit Theorem

Let X_i $(i = 1, 2, \cdots, n) \sim p_X(\mu, \sigma^2)$ with mean μ and variance σ^2 known.

Central Limit Theorem

Let X_i $(i = 1, 2, \cdots, n) \sim p_X(\mu, \sigma^2)$ with mean μ and variance σ^2 known.

Sample mean of these n draws: $S_n = \sum_{i=1}^{n} X_i/n$. From Bienaymé's Identity, $\sigma(S_n) = \dfrac{\sigma}{\sqrt{n}}$.

Standardise: define the random variable $Z_n = \dfrac{S_n - \mu}{\sigma/\sqrt{n}}$

Central Limit Theorem

Let X_i $(i = 1, 2, \cdots, n) \sim p_X(\mu, \sigma^2)$ with mean μ and variance σ^2 known.

Sample mean of these n draws: $S_n = \sum_{i=1}^{n} X_i/n$. From Bienaymé's Identity, $\sigma(S_n) = \dfrac{\sigma}{\sqrt{n}}$.

Standardise: define the random variable $Z_n = \dfrac{S_n - \mu}{\sigma/\sqrt{n}}$

The Central Limit Theorem can be stated in terms of the CDF of Z:

$$\lim_{n\to\infty} P\left(\frac{S_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \lim_{n\to\infty} P(Z \leq z) = \Phi(z).$$

Central Limit Theorem

Let X_i $(i = 1, 2, \cdots, n) \sim p_X(\mu, \sigma^2)$ with mean μ and variance σ^2 known.

Sample mean of these n draws: $S_n = \sum_{i=1}^{n} X_i/n$. From Bienaymé's Identity, $\sigma(S_n) = \dfrac{\sigma}{\sqrt{n}}$.

Standardise: define the random variable $Z_n = \dfrac{S_n - \mu}{\sigma/\sqrt{n}}$

The Central Limit Theorem can be stated in terms of the CDF of Z:

$$\lim_{n\to\infty} P\left(\frac{S_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \lim_{n\to\infty} P(Z \leq z) = \Phi(z).$$

Central Limit Theorem

Let X_i $(i = 1, 2, \cdots, n) \sim p_X(\mu, \sigma^2)$ with mean μ and variance σ^2 known.

Sample mean of these n draws: $S_n = \sum_{i=1}^{n} X_i/n$. From Bienaymé's Identity, $\sigma(S_n) = \frac{\sigma}{\sqrt{n}}$.

Standardise: define the random variable $Z_n = \frac{S_n - \mu}{\sigma/\sqrt{n}}$

The Central Limit Theorem can be stated in terms of the CDF of Z:

$$\lim_{n\to\infty} P\left(\frac{S_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \lim_{n\to\infty} P(Z \leq z) = \Phi(z).$$

For any PDF with finite variance, the sample mean of n random deviates is approximately normally distributed about the mean of the PDF, with variance $1/\sqrt{n}$ times the variance of the PDF.

Central Limit Theorem

Let X_i $(i = 1, 2, \cdots, n) \sim p_X(\mu, \sigma^2)$ with mean μ and variance σ^2 known.

Sample mean of these n draws: $S_n = \sum_{i=1}^{n} X_i / n$. From Bienaymé's Identity, $\sigma(S_n) = \dfrac{\sigma}{\sqrt{n}}$.

Standardise: define the random variable $Z_n = \dfrac{S_n - \mu}{\sigma/\sqrt{n}}$

The Central Limit Theorem can be stated in terms of the CDF of Z:

$$\lim_{n\to\infty} P\left(\frac{S_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \lim_{n\to\infty} P(Z \leq z) = \Phi(z).$$

For any PDF with finite variance, the sample mean of n random deviates is approximately normally distributed about the mean of the PDF, with variance $1/\sqrt{n}$ times the variance of the PDF.

One reason the Normal distribution pops up everywhere.

Aside: Maximising information entropy

The Shannon information entropy for continuous distributions, $H = -\int dx\ p_X(x) \log p_X(x)$, is the average level of information provided by a PDF.

For a given set of constraints, the PDF that maximises the entropy is the one that provides the least information about the system.

The distribution that maximises H subject to the normalisation constraint is the Uniform distribution (derivation).

If, in addition, we know the mean, the Exponential distribution is the maximum-entropy distribution.

If, in addition, we also know the variance, the maximum-entropy distribution is the Normal distribution.

For more information, read this article.

Functions of random variables

Probability Integral Transform

CDF of a random variable X: $F_X(x) \equiv P(X \leq x)$. Range: [0, 1] (like Standard Uniform).

- Monotonically $\uparrow \implies$ unique inverse (Quantile Function): $F_X^{-1}(q) = x$ such that $F_X(x) = q$.
- Range of $F_X(x)$: [0, 1], like the Standard Uniform distribution.
- CDF of the Standard Uniform distribution: $Y \sim \mathrm{Uniform}[0, 1] \implies F_Y(y) = y$.

Probability Integral Transform

CDF of a random variable X: $F_X(x) \equiv P(X \leq x)$. Range: [0, 1] (like Standard Uniform).

Monotonically $\uparrow \implies$ unique inverse (Quantile Function): $F_X^{-1}(q) = x$ such that $F_X(x) = q$.

Range of $F_X(x)$: [0, 1], like the Standard Uniform distribution.

CDF of the Standard Uniform distribution: $Y \sim \text{Uniform}[0, 1] \implies F_Y(y) = y$.

Probability Integral Transform:

If $Y = F_X(x)$ for some continuous random variable X, then $Y \sim \text{Uniform}[0, 1]$.

Proof:

Probability Integral Transform

CDF of a random variable X: $F_X(x) \equiv P(X \leq x)$. Range: [0, 1] (like Standard Uniform).

Monotonically $\uparrow \implies$ unique inverse (Quantile Function): $F_X^{-1}(q) = x$ such that $F_X(x) = q$.

Range of $F_X(x)$: [0, 1], like the Standard Uniform distribution.

CDF of the Standard Uniform distribution: $Y \sim \mathrm{Uniform}[0, 1] \implies F_Y(y) = y$.

Probability Integral Transform:

If $Y = F_X(x)$ for some continuous random variable X, then $Y \sim \mathrm{Uniform}[0, 1]$.

Proof:

$F_Y(y) \equiv P(Y \leq y) = P(F_X(x) \leq y)$

Probability Integral Transform

CDF of a random variable X: $F_X(x) \equiv P(X \leq x)$. Range: [0, 1] (like Standard Uniform).

Monotonically $\uparrow \implies$ unique inverse (Quantile Function): $F_X^{-1}(q) = x$ such that $F_X(x) = q$.

Range of $F_X(x)$: [0, 1], like the Standard Uniform distribution.

CDF of the Standard Uniform distribution: $Y \sim \mathrm{Uniform}[0, 1] \implies F_Y(y) = y$.

Probability Integral Transform:

If $Y = F_X(x)$ for some continuous random variable X, then $Y \sim \mathrm{Uniform}[0, 1]$.

Proof:

$$F_Y(y) \equiv P(Y \leq y) = P(F_X(x) \leq y)$$
$$= P(X \leq F_X^{-1}(y))$$

Probability Integral Transform

CDF of a random variable X: $F_X(x) \equiv P(X \leq x)$. Range: [0, 1] (like Standard Uniform).

Monotonically $\uparrow \implies$ unique inverse (Quantile Function): $F_X^{-1}(q) = x$ such that $F_X(x) = q$.

Range of $F_X(x)$: [0, 1], like the Standard Uniform distribution.

CDF of the Standard Uniform distribution: $Y \sim \mathrm{Uniform}[0,1] \implies F_Y(y) = y$.

Probability Integral Transform:

If $Y = F_X(x)$ for some continuous random variable X, then $Y \sim \mathrm{Uniform}[0,1]$.

Proof:

$$F_Y(y) \equiv P(Y \leq y) = P(F_X(x) \leq y)$$
$$= P(X \leq F_X^{-1}(y)) \quad \equiv F_X(F_X^{-1}(y)) = y \text{ (the CDF for } \mathrm{Uniform}[0,1]\text{!)}$$

Probability Integral Transform

CDF of a random variable X: $F_X(x) \equiv P(X \leq x)$. Range: [0, 1] (like Standard Uniform).

Monotonically $\uparrow \implies$ unique inverse (Quantile Function): $F_X^{-1}(q) = x$ such that $F_X(x) = q$.

Range of $F_X(x)$: [0, 1], like the Standard Uniform distribution.

CDF of the Standard Uniform distribution: $Y \sim \mathrm{Uniform}[0,1] \implies F_Y(y) = y$.

Probability Integral Transform:

If $Y = F_X(x)$ for some continuous random variable X, then $Y \sim \mathrm{Uniform}[0,1]$.

Proof:

$$F_Y(y) \equiv P(Y \leq y) = P(F_X(x) \leq y)$$
$$= P(X \leq F_X^{-1}(y)) \quad \equiv F_X(F_X^{-1}(y)) = y \text{ (the CDF for Uniform[0, 1]!)}$$

Example: Draw random variates from the distribution $p_X(x) = \frac{2}{\pi}\frac{1}{x^2+1}, \quad 0 \leq x < \infty$

Probability Integral Transform

CDF of a random variable X: $F_X(x) \equiv P(X \leq x)$. Range: [0, 1] (like Standard Uniform).

Monotonically $\uparrow \implies$ unique inverse (Quantile Function): $F_X^{-1}(q) = x$ such that $F_X(x) = q$.

Range of $F_X(x)$: [0, 1], like the Standard Uniform distribution.

CDF of the Standard Uniform distribution: $Y \sim \mathrm{Uniform}[0,1] \implies F_Y(y) = y$.

Probability Integral Transform:

If $Y = F_X(x)$ for some continuous random variable X, then $Y \sim \mathrm{Uniform}[0,1]$.

Proof:

$$F_Y(y) \equiv P(Y \leq y) = P(F_X(x) \leq y)$$
$$= P(X \leq F_X^{-1}(y)) \quad \equiv F_X(F_X^{-1}(y)) = y \text{ (the CDF for Uniform[0, 1]!)}$$

Example: Draw random variates from the distribution $p_X(x) = \frac{2}{\pi} \frac{1}{x^2 + 1}, \quad 0 \leq x < \infty$

Step 1: Compute the CDF: $F_X(x) \equiv P(X \leq x) = \int_0^x \frac{2}{\pi} \frac{1}{t^2 + 1} \, dt = \tan^{-1} x$.

Probability Integral Transform

CDF of a random variable X: $F_X(x) \equiv P(X \leq x)$. Range: [0, 1] (like Standard Uniform).

Monotonically $\uparrow \implies$ unique inverse (Quantile Function): $F_X^{-1}(q) = x$ such that $F_X(x) = q$.

Range of $F_X(x)$: [0, 1], like the Standard Uniform distribution.

CDF of the Standard Uniform distribution: $Y \sim \mathrm{Uniform}[0,1] \implies F_Y(y) = y$.

Probability Integral Transform:

If $Y = F_X(x)$ for some continuous random variable X, then $Y \sim \mathrm{Uniform}[0,1]$.

Proof:

$$F_Y(y) \equiv P(Y \leq y) = P(F_X(x) \leq y)$$
$$= P(X \leq F_X^{-1}(y)) \quad \equiv F_X(F_X^{-1}(y)) = y \text{ (the CDF for Uniform}[0,1]!)$$

Example: Draw random variates from the distribution $p_X(x) = \frac{2}{\pi}\frac{1}{x^2+1}, \quad 0 \leq x < \infty$

Step 1: Compute the CDF: $F_X(x) \equiv P(X \leq x) = \int_0^x \frac{2}{\pi}\frac{1}{t^2+1}\, dt = \tan^{-1} x$.

Step 2: Define $Y = F_X(x) = \tan^{-1} X$ ($\implies X = \tan Y$) $\implies Y \sim \mathrm{Uniform}[0,1]$.

Probability Integral Transform

CDF of a random variable X: $F_X(x) \equiv P(X \leq x)$. Range: [0, 1] (like Standard Uniform).

Monotonically $\uparrow \implies$ unique inverse (Quantile Function): $F_X^{-1}(q) = x$ such that $F_X(x) = q$.

Range of $F_X(x)$: [0, 1], like the Standard Uniform distribution.

CDF of the Standard Uniform distribution: $Y \sim \mathrm{Uniform}[0,1] \implies F_Y(y) = y$.

Probability Integral Transform:

If $Y = F_X(x)$ for some continuous random variable X, then $Y \sim \mathrm{Uniform}[0,1]$.

Proof:

$$F_Y(y) \equiv P(Y \leq y) = P(F_X(x) \leq y)$$
$$= P(X \leq F_X^{-1}(y)) \quad \equiv F_X(F_X^{-1}(y)) = y \text{ (the CDF for } \mathrm{Uniform}[0,1]\text{!)}$$

Example: Draw random variates from the distribution $p_X(x) = \frac{2}{\pi}\frac{1}{x^2+1}, \quad 0 \leq x < \infty$

Step 1: Compute the CDF: $F_X(x) \equiv P(X \leq x) = \int_0^x \frac{2}{\pi}\frac{1}{t^2+1}\, dt = \tan^{-1} x$.

Step 2: Define $Y = F_X(x) = \tan^{-1} X$ ($\implies X = \tan Y$) $\implies Y \sim \mathrm{Uniform}[0,1]$.

Step 3: Draw samples y from $\mathrm{Uniform}[0,1]$ and then compute $x = \tan y$.

This procedure guarantees that the resulting x will be distributed according to $p_X(x)$.

PDFs of functions of random variables

A function of a random variable will, in general, not have the same distribution.

In simple cases, it is possible to analytically determine the PDF of the function.

This lecture: 3 analytical methods for determining the PDFs of functions of random variables.

PDFs of functions of random variables

A function of a random variable will, in general, not have the same distribution.

In simple cases, it is possible to analytically determine the PDF of the function.

This lecture: 3 analytical methods for determining the PDFs of functions of random variables.

For more complicated functions, or when we don't even know the functional dependence analytically, we have to resort to numerical methods.

One common example where such sampling is required is in Monte Carlo methods, especially when computing and sampling the posterior probability distributions in Bayesian analysis.

These methods allow us to compute distributions for complicated functions by sampling simple distributions such $\mathrm{Uniform}[0,1]$.

PDFs of functions of random variables

A function of a random variable will, in general, not have the same distribution.

In simple cases, it is possible to analytically determine the PDF of the function.

This lecture: 3 analytical methods for determining the PDFs of functions of random variables.

For more complicated functions, or when we don't even know the functional dependence analytically, we have to resort to numerical methods.

One common example where such sampling is required is in Monte Carlo methods, especially when computing and sampling the posterior probability distributions in Bayesian analysis.

These methods allow us to compute distributions for complicated functions by sampling simple distributions such $\mathrm{Uniform}[0,1]$.

Examples of analytical cases to which such analysis is applicable:

1D general case: If $X \sim p_X(x)$ and $Y = f(X)$, what is $p_Y(y)$?

2D independent case: If $X \sim p_X(x)$, $Y \sim p_Y(y)$ and $Z = f(X, Y)$, what is $p_Z(z)$?

2D general case: If $X, Y \sim p_{X,Y}(x, y)$ and $Z = f(X, Y)$, what is $p_Z(z)$?

2D linear combination: If $X, Y \sim p_{X,Y}(x, y)$ and $Z = \alpha X + \beta Y$, what is $p_Z(z)$?

(and higher dimension equivalents).

PDFs of functions: CDF method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

PDFs of functions: CDF method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

Step 1: Using allowed range for X, find allowed range of Y.

Step 2: Find CDF of Y over this range.

Step 3: If CDF is continuous, differentiate w.r.t. y to obtain PDF of Y over the range.

PDFs of functions: CDF method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

Step 1: Using allowed range for X, find allowed range of Y.

Step 2: Find CDF of Y over this range.

Step 3: If CDF is continuous, differentiate w.r.t. y to obtain PDF of Y over the range.

Example: If $X \sim \mathrm{Uniform}[-1, 1]$ and $Y = X^2$, find $p_Y(y)$.

PDFs of functions: CDF method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

Step 1: Using allowed range for X, find allowed range of Y.

Step 2: Find CDF of Y over this range.

Step 3: If CDF is continuous, differentiate w.r.t. y to obtain PDF of Y over the range.

Example: If $X \sim \mathrm{Uniform}[-1, 1]$ and $Y = X^2$, find $p_Y(y)$.

Note that the inverse of $f(X)$ in this case is multi-valued: $y = x^2 \Longrightarrow x = \pm\sqrt{y}$.

PDFs of functions: CDF method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

Step 1: Using allowed range for X, find allowed range of Y.

Step 2: Find CDF of Y over this range.

Step 3: If CDF is continuous, differentiate w.r.t. y to obtain PDF of Y over the range.

Example: If $X \sim \mathrm{Uniform}[-1, 1]$ and $Y = X^2$, find $p_Y(y)$.

Note that the inverse of $f(X)$ in this case is multi-valued: $y = x^2 \Longrightarrow x = \pm\sqrt{y}$.

CDF of X: $F_X(x) = \dfrac{x - (-1)}{1 - (-1)} = \dfrac{x+1}{2}, \qquad x \in [-1, 1]$.

PDFs of functions: CDF method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

Step 1: Using allowed range for X, find allowed range of Y.

Step 2: Find CDF of Y over this range.

Step 3: If CDF is continuous, differentiate w.r.t. y to obtain PDF of Y over the range.

Example: If $X \sim \mathrm{Uniform}[-1, 1]$ and $Y = X^2$, find $p_Y(y)$.

Note that the inverse of $f(X)$ in this case is multi-valued: $y = x^2 \implies x = \pm\sqrt{y}$.

CDF of X: $F_X(x) = \dfrac{x-(-1)}{1-(-1)} = \dfrac{x+1}{2}, \qquad x \in [-1, 1]$.

Step 1: If $x \in [-1, 1], y = x^2 \in [0, 1]$.

Step 2: CDF of Y: $F_Y(y) \equiv P(Y \leq y) = P(-\sqrt{y} \leq x \leq \sqrt{y})$

$$= P(-1 \leq x \leq \sqrt{y}) - P(-1 \leq x \leq -\sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \frac{\sqrt{y}+1}{2} - \frac{1-\sqrt{y}}{2} = \sqrt{y}, \text{ for } 0 \leq y \leq 1.$$

Step 3: Since the CDF is continuous, $p_Y(y) = \dfrac{d}{dy} F_Y(y) = \dfrac{d}{dy}\sqrt{y} = \dfrac{1}{2\sqrt{y}}$.

PDFs of functions: Transformation (Inverse) method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

This method works if inverse can be written down explicitly. (not the case with, *e.g.*, $y = xe^x$)

$Y = f(X) \Longrightarrow X = f^{-1}(Y)$. Write x in terms of y everywhere.

$(x, x + dx) \longrightarrow (y, y + dy)$, such that $p_X(x)\ dx = p_Y(y)\ dy \Longrightarrow p_Y(y) = \dfrac{p_X(f^{-1}(y))}{\left|\dfrac{dy}{dx}\right|}$.

PDFs of functions: Transformation (Inverse) method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

This method works if inverse can be written down explicitly. (not the case with, *e.g.*, $y = xe^x$)

$Y = f(X) \implies X = f^{-1}(Y)$. Write x in terms of y everywhere.

$(x, x+dx) \longrightarrow (y, y+dy)$, such that $p_X(x)\ dx = p_Y(y)\ dy \implies p_Y(y) = \dfrac{p_X(f^{-1}(y))}{\left|\dfrac{dy}{dx}\right|}$.

If inverse multi-valued, add contributions to $p_Y(y)$: $p_Y(y) = \sum_{x_i} \dfrac{p_X(x_i)}{\left|\dfrac{dy}{dx}\right|} \quad \forall\ x_i : f(x_i) = y$

PDFs of functions: Transformation (Inverse) method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

This method works if inverse can be written down explicitly. (not the case with, *e.g.*, $y = xe^x$)

$Y = f(X) \implies X = f^{-1}(Y)$. Write x in terms of y everywhere.

$(x, x + dx) \longrightarrow (y, y + dy)$, such that $p_X(x)\ dx = p_Y(y)\ dy \implies p_Y(y) = \dfrac{p_X(f^{-1}(y))}{\left|\dfrac{dy}{dx}\right|}$.

If inverse multi-valued, add contributions to $p_Y(y)$: $p_Y(y) = \displaystyle\sum_{x_i} \dfrac{p_X(x_i)}{\left|\dfrac{dy}{dx}\right|} \quad \forall\ x_i : f(x_i) = y$

Example: $X \sim \mathcal{N}(0,1) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $Y = X^2$

PDFs of functions: Transformation (Inverse) method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

This method works if inverse can be written down explicitly. (not the case with, *e.g.*, $y = xe^x$)

$Y = f(X) \implies X = f^{-1}(Y)$. Write x in terms of y everywhere.

$(x, x + dx) \longrightarrow (y, y + dy)$, such that $p_X(x)\ dx = p_Y(y)\ dy \implies p_Y(y) = \dfrac{p_X(f^{-1}(y))}{\left|\dfrac{dy}{dx}\right|}$.

If inverse multi-valued, add contributions to $p_Y(y)$: $p_Y(y) = \displaystyle\sum_{x_i} \frac{p_X(x_i)}{\left|\dfrac{dy}{dx}\right|} \quad \forall\ x_i : f(x_i) = y$

Example: $X \sim \mathcal{N}(0, 1) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $Y = X^2$

$x \in (-\infty, \infty) \implies y \in (0, \infty)$ (0 should be included, but see PDF below!).

PDFs of functions: Transformation (Inverse) method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

This method works if inverse can be written down explicitly. (not the case with, *e.g.*, $y = xe^x$)

$Y = f(X) \implies X = f^{-1}(Y)$. Write x in terms of y everywhere.

$(x, x + dx) \longrightarrow (y, y + dy)$, such that $p_X(x)\ dx = p_Y(y)\ dy \implies p_Y(y) = \dfrac{p_X(f^{-1}(y))}{\left|\dfrac{dy}{dx}\right|}$.

If inverse multi-valued, add contributions to $p_Y(y)$: $p_Y(y) = \displaystyle\sum_{x_i} \frac{p_X(x_i)}{\left|\dfrac{dy}{dx}\right|} \quad \forall\ x_i : f(x_i) = y$

Example: $X \sim \mathcal{N}(0,1) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $Y = X^2$

$x \in (-\infty, \infty) \implies y \in (0, \infty)$ (0 should be included, but see PDF below!).

Invert: $x = \pm\sqrt{y}$ – multivalued but symmetric (equal contribution to PDF from each root).

Write x in terms of y everywhere: $p_X(x) = p_X(f^{-1}(y)) = \dfrac{1}{\sqrt{2\pi}} e^{-y/2}$, $\left|\dfrac{dy}{dx}\right| = 2\sqrt{y}$.

PDFs of functions: Transformation (Inverse) method

Aim: Given $X \sim p_X(x)$ and $Y = f(X)$, find $p_Y(y)$.

This method works if inverse can be written down explicitly. (not the case with, *e.g.*, $y = xe^x$)

$Y = f(X) \implies X = f^{-1}(Y)$. Write x in terms of y everywhere.

$(x, x + dx) \longrightarrow (y, y + dy)$, such that $p_X(x)\ dx = p_Y(y)\ dy \implies p_Y(y) = \dfrac{p_X(f^{-1}(y))}{\left|\dfrac{dy}{dx}\right|}$.

If inverse multi-valued, add contributions to $p_Y(y)$: $p_Y(y) = \displaystyle\sum_{x_i} \frac{p_X(x_i)}{\left|\dfrac{dy}{dx}\right|} \quad \forall\ x_i : f(x_i) = y$

Example: $X \sim \mathcal{N}(0,1) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}, Y = X^2$

$x \in (-\infty, \infty) \implies y \in (0, \infty)$ (0 should be included, but see PDF below!).

Invert: $x = \pm\sqrt{y}$ – multivalued but symmetric (equal contribution to PDF from each root).

Write x in terms of y everywhere: $p_X(x) = p_X(f^{-1}(y)) = \dfrac{1}{\sqrt{2\pi}} e^{-y/2}, \left|\dfrac{dy}{dx}\right| = 2\sqrt{y}$.

Apply multi-valued version of the transformation method. Factor of 2 in derivative cancels.

$\implies p_Y(y) = \dfrac{1}{\sqrt{2\pi}} \dfrac{e^{-y/2}}{y}$, which is the χ^2 distribution for 1 degree of freedom.

PDFs of linear combinations: convolution

If $Z = \alpha X + \beta Y$ for $\alpha, \beta \in \mathbb{R}$, $p_Z(z)$ can be written as a convolution of $p_X(x)$ and $p_Y(y)$.
(proof: see, *e.g.*, Slides 13-15 here.)

$$p_Z(z) = \int_{-\infty}^{\infty} dx\ p_X(x)\ p_Y\left(\frac{z - \alpha x}{\beta}\right)$$

(where y is written in terms of x and z)

PDFs of linear combinations: convolution

If $Z = \alpha X + \beta Y$ for $\alpha, \beta \in \mathbb{R}$, $p_Z(z)$ can be written as a convolution of $p_X(x)$ and $p_Y(y)$.
(proof: see, *e.g.*, Slides 13-15 here.)

$$p_Z(z) = \int_{-\infty}^{\infty} dx \; p_X(x) \; p_Y\left(\frac{z - \alpha x}{\beta}\right)$$

(where y is written in terms of x and z)

Example: sum of two uniform deviates

$X, Y \sim U[0, 1]$ and $Z = X + Y \implies Y = Z - X$.

PDFs of linear combinations: convolution

If $Z = \alpha X + \beta Y$ for $\alpha, \beta \in \mathbb{R}$, $p_Z(z)$ can be written as a convolution of $p_X(x)$ and $p_Y(y)$.
(proof: see, *e.g.*, Slides 13-15 here.)

$$p_Z(z) = \int_{-\infty}^{\infty} dx \; p_X(x) \; p_Y\left(\frac{z - \alpha x}{\beta}\right) \qquad \text{(where } y \text{ is written in terms of } x \text{ and } z\text{)}$$

Example: sum of two uniform deviates

$X, Y \sim U[0, 1]$ and $Z = X + Y \implies Y = Z - X$.

x and y go from 0 to 1, so z goes from 0 to 2. Split this interval into [0, 1] and [1, 2].

For $z \in (0, 1)$, we need $z - x \geq 0$, so $x < z$: $\int_0^z dx \; p_X(x) \; p_Y((z - x)) = \int_0^z dx = z$.

For $z \in (1, 2)$, we need $z - x \leq 1$, so $x > z - 1$: $\int_{z-1}^1 dx \; p_X(x) \; p_Y((z - x)) = \int_{z-1}^1 dx = 2 - z$.

PDFs of linear combinations: convolution

If $Z = \alpha X + \beta Y$ for $\alpha, \beta \in \mathbb{R}$, $p_Z(z)$ can be written as a convolution of $p_X(x)$ and $p_Y(y)$.
(proof: see, *e.g.*, Slides 13-15 here.)

$$p_Z(z) = \int_{-\infty}^{\infty} dx \; p_X(x) \; p_Y\left(\frac{z - \alpha x}{\beta}\right) \qquad \text{(where } y \text{ is written in terms of } x \text{ and } z\text{)}$$

Example: sum of two uniform deviates

$X, Y \sim U[0,1]$ and $Z = X + Y \implies Y = Z - X$.

x and y go from 0 to 1, so z goes from 0 to 2. Split this interval into [0, 1] and [1, 2].

For $z \in (0,1)$, we need $z - x \geq 0$, so $x < z$: $\int_0^z dx \; p_X(x) \; p_Y((z-x)) = \int_0^z dx = z$.

For $z \in (1,2)$, we need $z - x \leq 1$, so $x > z - 1$: $\int_{z-1}^1 dx \; p_X(x) \; p_Y((z-x)) = \int_{z-1}^1 dx = 2 - z$.

$$\implies p_Z(z) = \begin{cases} z & : 0 < z \leq 1 \\ 2 - z & : 1 < z \leq 2 \end{cases}$$

Can be extended to a linear combination of N random variables!