Statistics for Astronomers: Lecture 8, 2020.10.26

Prof. Sundar Srinivasan

IRyA/UNAM

Review

The χ^2 distribution.

χ^2 minimisation and interpretation. Reduced χ^2 and caution.

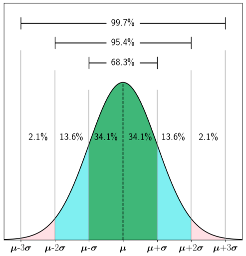
The Empirical Rule for normal distributions.

The z-score.

The Empirical Rule for normal distributions

Probability of obtaining values within 1, 2, or 3 σ of the centre of a normal distribution.

Central probability	Extreme (right-tail) probability
$P(\lvert Z\rvert \leq 1) \approx 0.68$	$P(Z > 1) = \frac{1}{2}\left[1 - P(\lvert Z\rvert \leq 1)\right] \approx 0.16$
$P(\lvert Z\rvert \leq 2) \approx 0.95$	$P(Z > 2) \approx 0.025.$
$P(\lvert Z\rvert \leq 3) \approx 0.997$	$P(Z > 3) \approx 0.0015.$



Code for plot available here.

Therefore also known as the 68–95–99.7 Rule.

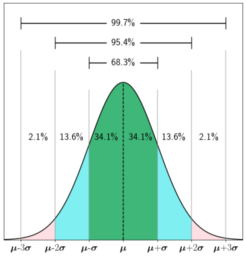
3σ rule of thumb for normal distributions:

most (99.7%) of your data is within 3σ of the mean.

The Empirical Rule for normal distributions

Probability of obtaining values within 1, 2, or 3 σ of the centre of a normal distribution.

Central probability	Extreme (right-tail) probability
$P(\|Z\| \leq 1) \approx 0.68$	$P(Z > 1) = \frac{1}{2}\left[1 - P(\|Z\| \leq 1)\right] \approx 0.16$
$P(\|Z\| \leq 2) \approx 0.95$	$P(Z > 2) \approx 0.025.$
$P(\|Z\| \leq 3) \approx 0.997$	$P(Z > 3) \approx 0.0015.$



Code for plot available here.

Therefore also known as the 68–95–99.7 Rule.

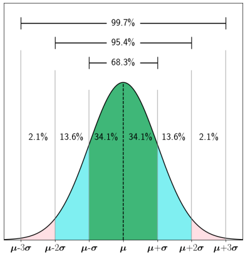
3σ rule of thumb for normal distributions:
most (99.7%) of your data is within 3σ of the mean.

Detecting new particle: min 5σ. $P(Z > 5) \approx 5.7 \times 10^{-7}$
Highly unlikely that event due to random chance.

The Empirical Rule for normal distributions

Probability of obtaining values within 1, 2, or 3 σ of the centre of a normal distribution.

Central probability	Extreme (right-tail) probability
$P(\lvert Z\rvert \le 1) \approx 0.68$	$P(Z > 1) = \frac{1}{2}\left[1 - P(\lvert Z\rvert \le 1)\right] \approx 0.16$
$P(\lvert Z\rvert \le 2) \approx 0.95$	$P(Z > 2) \approx 0.025.$
$P(\lvert Z\rvert \le 3) \approx 0.997$	$P(Z > 3) \approx 0.0015.$



Code for plot available here.

Therefore also known as the 68–95–99.7 Rule.

3σ rule of thumb for normal distributions:
most (99.7%) of your data is within 3σ of the mean.

Detecting new particle: min 5σ. $P(Z > 5) \approx 5.7 \times 10^{-7}$
Highly unlikely that event due to random chance.

Other way around: for significance level $1 - \alpha$,

what is the threshold $z_{\alpha/2}$ such that $P(|Z| > z_{\alpha/2}) < \alpha$? (Two-tailed test)

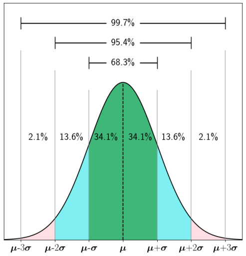
what is the threshold $z_{\alpha/2}$ such that $P(Z > z_{\alpha/2}) < \alpha$? (One-tailed test)

$$P(|Z| > z_{\alpha/2}) = 1 - P(|Z| \le z_{\alpha/2}) = 1 - \mathrm{erf}\left(\frac{z_{\alpha/2}}{\sqrt{2}}\right) \implies z_{\alpha/2} = \sqrt{2}\,\mathrm{erf}^{-1}\left(1 - \alpha\right).$$

The Empirical Rule for normal distributions

Probability of obtaining values within 1, 2, or 3 σ of the centre of a normal distribution.

Central probability	Extreme (right-tail) probability
$P(\|Z\| \le 1) \approx 0.68$	$P(Z > 1) = \frac{1}{2}\left[1 - P(\|Z\| \le 1)\right] \approx 0.16$
$P(\|Z\| \le 2) \approx 0.95$	$P(Z > 2) \approx 0.025.$
$P(\|Z\| \le 3) \approx 0.997$	$P(Z > 3) \approx 0.0015.$



Code for plot available here.

Therefore also known as the 68–95–99.7 Rule.

3σ rule of thumb for normal distributions:
most (99.7%) of your data is within 3σ of the mean.

Detecting new particle: min 5σ. $P(Z > 5) \approx 5.7 \times 10^{-7}$
Highly unlikely that event due to random chance.

Other way around: for significance level $1 - \alpha$,
what is the threshold $z_{\alpha/2}$ such that $P(|Z| > z_{\alpha/2}) < \alpha$? (Two-tailed test)
what is the threshold $z_{\alpha/2}$ such that $P(Z > z_{\alpha/2}) < \alpha$? (One-tailed test)

$$P(|Z| > z_{\alpha/2}) = 1 - P(|Z| \le z_{\alpha/2}) = 1 - \mathrm{erf}\left(\frac{z_{\alpha/2}}{\sqrt{2}}\right) \implies z_{\alpha/2} = \sqrt{2}\,\mathrm{erf}^{-1}\left(1 - \alpha\right).$$

Hypothesis testing. $1 - \alpha$ usually 95%. Observed probability: p-value.

Example: pixel with flux 3σ above noise level. p-value: $P(Z > 3) = 0.00135 < \alpha = 0.05$.
$\implies$ the detection is statistically significant at the $\alpha = 95\%$ level.

Student's t-distribution

If $X \sim \mathcal{N}(\mu, \sigma^2)$ with μ unknown, we use $\overline{X}$ and σ to estimate μ: $Z = \dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$.

Student's t-distribution

If $X \sim \mathcal{N}(\mu, \sigma^2)$ with μ unknown, we use $\overline{X}$ and σ to estimate μ: $Z = \dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$.

What if σ is also unknown? We can estimate it from data. Recall: $\widehat{\sigma} = S$ such that

$$S^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \quad (\mu \text{ known})$$

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2 \quad (\mu \text{ unknown})$$

Student's t-distribution

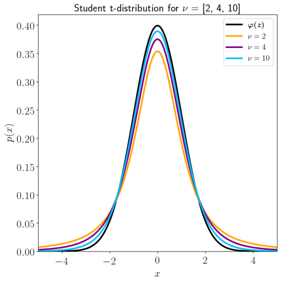
If $X \sim \mathcal{N}(\mu, \sigma^2)$ with μ unknown, we use $\overline{X}$ and σ to estimate μ: $Z = \dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)$.

What if σ is also unknown? We can estimate it from data. Recall: $\widehat{\sigma} = S$ such that

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \quad (\mu \text{ known}) \qquad S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2 \quad (\mu \text{ unknown})$$

$T \equiv \dfrac{X - \mu}{S}$ has a Student's t-distribution with $\#\mathrm{dof} = N - 1$ (if μ estimated by $\overline{x}$).

$$p_T(t, \nu) \propto \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$ with $\nu = \#\mathrm{dof} = N$ or $N - 1$.



Code for plot available here.

Student's t-distribution

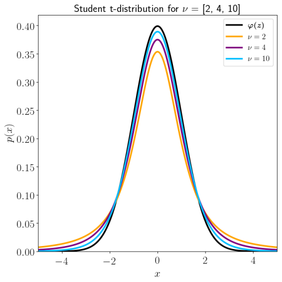
If $X \sim \mathcal{N}(\mu, \sigma^2)$ with μ unknown, we use $\overline{X}$ and σ to estimate μ: $Z = \dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)$.

What if σ is also unknown? We can estimate it from data. Recall: $\widehat{\sigma} = S$ such that

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \quad (\mu \text{ known}) \qquad\qquad S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2 \quad (\mu \text{ unknown})$$

$T \equiv \dfrac{X - \mu}{S}$ has a Student's t-distribution with $\#\text{dof} = N - 1$ (if μ estimated by $\overline{x}$).

$$p_T(t, \nu) \propto \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} \text{ with } \nu = \#\text{dof} = N \text{ or } N - 1.$$



Code for plot available here.

Symmetric about $t = 0$, odd moments $= 0$ (like the Gaussian).
Uncertain estimate S for $\sigma \Rightarrow$ more probability in the tails.

$$p_T(t, \nu) \xrightarrow{N \to \infty} \mathcal{N}(0,1) \qquad\qquad \mathrm{Var}[T] = \sqrt{\frac{\nu}{\nu - 2}} \xrightarrow{\nu \to \infty} 1.$$

Student's t-distribution

If $X \sim \mathcal{N}(\mu, \sigma^2)$ with μ unknown, we use $\overline{X}$ and σ to estimate μ: $Z = \dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$.

What if σ is also unknown? We can estimate it from data. Recall: $\widehat{\sigma} = S$ such that

$$S^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \quad (\mu \text{ known}) \qquad S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2 \quad (\mu \text{ unknown})$$

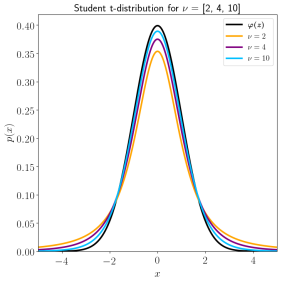
$T \equiv \dfrac{X - \mu}{S}$ has a Student's t-distribution with #dof $= N - 1$ (if μ estimated by $\overline{x}$).

$$p_T(t, \nu) \propto \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} \text{ with } \nu = \#\text{dof} = N \text{ or } N - 1.$$

Symmetric about $t = 0$, odd moments $= 0$ (like the Gaussian).
Uncertain estimate S for $\sigma \Rightarrow$ more probability in the tails.

$$p_T(t, \nu) \xrightarrow{N \to \infty} \mathcal{N}(0, 1) \qquad \mathrm{Var}[T] = \sqrt{\frac{\nu}{\nu - 2}} \xrightarrow{\nu \to \infty} 1.$$

Use the t-distribution for $N < 30$.



Code for plot available here.

The t statistic

For small samples ($N < 30$), we must compute the t-equivalent of the z statistic in order to determine t-scores.

For a Normal random variable X, $T = \dfrac{X - \mu}{S}$. For any (CLT) sample mean: $T = \dfrac{\overline{X} - \mu}{S/\sqrt{N}}$

The t statistic

For small samples ($N < 30$), we must compute the t-equivalent of the z statistic in order to determine t-scores.

For a Normal random variable X, $T = \dfrac{X - \mu}{S}$. For any (CLT) sample mean: $T = \dfrac{\overline{X} - \mu}{S/\sqrt{N}}$

For $\nu = 4$, using `scipy.stats.t.cdf` and `scipy.stats.norm.cdf`,
compare the central concentration of T with Z:
$P(|T_{\nu=4}| < 1) \approx 0.63$; $P(|Z| < 1) \approx 0.68$
$P(|T_{\nu=4}| < 2) \approx 0.88$; $P(|Z| < 2) \approx 0.95$
$P(|T_{\nu=4}| < 3) \approx 0.96$; $P(|Z| < 3) \approx 0.997$

The t statistic

For small samples ($N < 30$), we must compute the t-equivalent of the z statistic in order to determine t-scores.

For a Normal random variable X, $T = \dfrac{X - \mu}{S}$. For any (CLT) sample mean: $T = \dfrac{\overline{X} - \mu}{S/\sqrt{N}}$

For $\nu = 4$, using `scipy.stats.t.cdf` and `scipy.stats.norm.cdf`, compare the central concentration of T with Z:

$P(|T_{\nu=4}| < 1) \approx 0.63;\ P(|Z| < 1) \approx 0.68$

$P(|T_{\nu=4}| < 2) \approx 0.88;\ P(|Z| < 2) \approx 0.95$

$P(|T_{\nu=4}| < 3) \approx 0.96;\ P(|Z| < 3) \approx 0.997$

Central behaviour quite similar!

The t statistic

For small samples ($N < 30$), we must compute the t-equivalent of the z statistic in order to determine t-scores.

For a Normal random variable X, $T = \dfrac{X - \mu}{S}$. For any (CLT) sample mean: $T = \dfrac{\overline{X} - \mu}{S/\sqrt{N}}$

For $\nu = 4$, using `scipy.stats.t.cdf` and `scipy.stats.norm.cdf`,
compare the central concentration of T with Z:
$P(|T_{\nu=4}| < 1) \approx 0.63;\; P(|Z| < 1) \approx 0.68$
$P(|T_{\nu=4}| < 2) \approx 0.88;\; P(|Z| < 2) \approx 0.95$
$P(|T_{\nu=4}| < 3) \approx 0.96;\; P(|Z| < 3) \approx 0.997$
Central behaviour quite similar!

For $\nu = 4$, using `scipy.stats.t.ppf` and `scipy.stats.norm.ppf`,
compare probability in the tails for various significance levels; that is,
$P(|T_{\nu=4}| > t_{\nu=4,\alpha/2}) < \alpha$ vs. $P(|Z| > z_{\alpha/2}) < \alpha$:
$\alpha = 0.1$: $t_{\nu=4,\alpha/2} \approx 2.13;\; z_{\alpha/2} \approx 1.64$ using `t.ppf(1−α/2)`, `norm.ppf(1−α/2)`
$\alpha = 0.05$: $t_{\nu=4,\alpha/2} \approx 2.78;\; z_{\alpha/2} \approx 1.96$
$\alpha = 0.003$: $t_{\nu=4,\alpha/2} \approx 6.44;\; z_{\alpha/2} \approx 2.97$

The t statistic

For small samples ($N < 30$), we must compute the t-equivalent of the z statistic in order to determine t-scores.

For a Normal random variable X, $T = \dfrac{X - \mu}{S}$. For any (CLT) sample mean: $T = \dfrac{\overline{X} - \mu}{S/\sqrt{N}}$

For $\nu = 4$, using `scipy.stats.t.cdf` and `scipy.stats.norm.cdf`,
compare the central concentration of T with Z:
$P(|T_{\nu=4}| < 1) \approx 0.63$; $P(|Z| < 1) \approx 0.68$
$P(|T_{\nu=4}| < 2) \approx 0.88$; $P(|Z| < 2) \approx 0.95$
$P(|T_{\nu=4}| < 3) \approx 0.96$; $P(|Z| < 3) \approx 0.997$
Central behaviour quite similar!

For $\nu = 4$, using `scipy.stats.t.ppf` and `scipy.stats.norm.ppf`,
compare probability in the tails for various significance levels; that is,
$P(|T_{\nu=4}| > t_{\nu=4,\alpha/2}) < \alpha$ vs. $P(|Z| > z_{\alpha/2}) < \alpha$:
$\alpha = 0.1$: $t_{\nu=4,\alpha/2} \approx 2.13$; $z_{\alpha/2} \approx 1.64$ using `t.ppf(1−α/2), norm.ppf(1−α/2)`
$\alpha = 0.05$: $t_{\nu=4,\alpha/2} \approx 2.78$; $z_{\alpha/2} \approx 1.96$
$\alpha = 0.003$: $t_{\nu=4,\alpha/2} \approx 6.44$; $z_{\alpha/2} \approx 2.97$

Very different because of behaviour in the tails ("fatter" or "heavier"-tailed distribution)!

More general rule(s) for non-Normal distributions?

If location and scale parameters known:

Definition (Chebyshev's Inequality)

If X is a random variable with finite mean μ and finite non-zero standard deviation σ, then

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geq k\right) \leq \frac{1}{k^2} \qquad \text{(valid for } k > 1\text{)},$$

Two-tailed version, can be modified for asymmetric distributions.

Ex: $P(|Z| \geq 2) \leq 0.25$; $P(|Z| \geq 3) \leq 0.11$; compare to Empirical Rule for Normal distributions.

More general rule(s) for non-Normal distributions?

If location and scale parameters known:

Definition (Chebyshev's Inequality)

If X is a random variable with finite mean μ and finite non-zero standard deviation σ, then

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geq k\right) \leq \frac{1}{k^2} \qquad \text{(valid for } k > 1\text{)},$$

Two-tailed version, can be modified for asymmetric distributions.
Ex: $P(|Z| \geq 2) \leq 0.25$; $P(|Z| \geq 3) \leq 0.11$; compare to Empirical Rule for Normal distributions.

If μ, σ unknown:

Definition (Markov's Inequality)

If X is a nonnegative random variable and $a > 0$,

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Two-tailed version also exists.

Interval estimates

Summary and references

When generating estimates for parameters, a point estimate alone is not enough.

Summary and references

When generating estimates for parameters, a point estimate alone is not enough.

An interval estimate is a range of values around the point estimate such that a probabilistic statement can be made about its relation to the true parameter value. Such a statement is usually in terms of the confidence we have that the range includes a certain fraction of possible values that are observed for the parameter in terms of its estimate.

Summary and references

When generating estimates for parameters, a point estimate alone is not enough.

An interval estimate is a range of values around the point estimate such that a probabilistic statement can be made about its relation to the true parameter value. Such a statement is usually in terms of the confidence we have that the range includes a certain fraction of possible values that are observed for the parameter in terms of its estimate.

Example: an experiment to determine the mass of a rock results in a measurement of (0.2 ± 0.05) kg, where 0.05 kg is the 1σ uncertainty in the measured mass. For Gaussian uncertainties, this means that 68.3% of the measured masses will lie in the range $[0.15, 0.25]$ kg.

Summary and references

When generating estimates for parameters, a point estimate alone is not enough.

An interval estimate is a range of values around the point estimate such that a probabilistic statement can be made about its relation to the true parameter value. Such a statement is usually in terms of the confidence we have that the range includes a certain fraction of possible values that are observed for the parameter in terms of its estimate.

Example: an experiment to determine the mass of a rock results in a measurement of (0.2 ± 0.05) kg, where 0.05 kg is the 1σ uncertainty in the measured mass. For Gaussian uncertainties, this means that 68.3% of the measured masses will lie in the range $[0.15, 0.25]$ kg.

Note that an interval estimate requires an estimate of the variance of the distribution of observed values. We will look at various ways of estimating this variance as well.

Summary and references

When generating estimates for parameters, a point estimate alone is not enough.

An interval estimate is a range of values around the point estimate such that a probabilistic statement can be made about its relation to the true parameter value. Such a statement is usually in terms of the confidence we have that the range includes a certain fraction of possible values that are observed for the parameter in terms of its estimate.

Example: an experiment to determine the mass of a rock results in a measurement of (0.2 ± 0.05) kg, where 0.05 kg is the 1σ uncertainty in the measured mass. For Gaussian uncertainties, this means that 68.3% of the measured masses will lie in the range $[0.15, 0.25]$ kg.

Note that an interval estimate requires an estimate of the variance of the distribution of observed values. We will look at various ways of estimating this variance as well.

References:

"Statistics: A Guide and Reference to the Use of Statistical Methods in the Physical Sciences" - R. J. Barlow.

"Dos and don'ts of reduced chi-squares" - R. Andrae, 2010.

Wall & Jenkins.