TUTORIAL WEEK 3

Recap: How to estimate a parameter from data (two methods)

MM or MOM (method of moments, slide 31 module02):
IDEA: Equate theoretical moments with sample moments
1. Set $E(X)=\bar{X}$ (if more than one parameters, set $E(X^2)=\overline{X^2}$, etc., one equation per parameter)
2. Find an expression for E(X), etc. as a function of parameter/s of interest (e.g. theta)
3. Rearrange equation/s in step 1, solving for parameter/s
4. You now have an estimator/s for parameter/s in terms of data.

MLE (maximum likelihood estimation, slide 45 module02):
IDEA: Which value of a parameter maximises the probability of getting the observed data?
For independent data, joint pmf/pdf is product of individual pmfs/pdfs.
Value of parameter that maximises the log of joint pmf/pdf will also maximise joint pmf/pdf (and often easier to use work with).
1. Write down the log-likelihood (or likelihood) expression (m=number of parameters)

$$\ell(\theta_1, \ldots, \theta_m) = \ln\left(L(\theta_1, \ldots, \theta_m)\right) = \ln\left(\prod_{i=1}^{n} f(x_i \mid \theta_1, \ldots, \theta_m)\right)$$

2. Consider if you can tell which value of parameter maximises the log-likelihood/likelihood
3. If not, differentiate the log-likelihood/likelihood (giving the 'score' equation/s) and set equal to 0

$$S(\theta_i) = \frac{\partial \ell}{\partial \theta_i} = 0$$

4. Rearrange equation/s in step 3 and solve for parameter/s of interest
5. You now have an estimator/s for parameter/s in terms of data.

Recap: Good properties of estimators

1. Unbiasedness

$$E(\hat{\theta}) = \theta$$

2. Small variance

TUTORIAL WEEK 4

$100 \cdot (1-\alpha)\%$ confidence interval

for the population mean μ $\boxed{(\sigma \text{ known})}$

$$\left(\bar{x} - c\frac{\sigma}{\sqrt{n}}, \bar{x} + c\frac{\sigma}{\sqrt{n}}\right)$$

where $\Phi^{-1}(1-\alpha/2) = c$

$100 \cdot (1-\alpha)\%$ confidence interval

for the population mean μ $\boxed{(\sigma \text{ unknown})}$

$$\left(\bar{x} - c\frac{s}{\sqrt{n}}, \bar{x} + c\frac{s}{\sqrt{n}}\right)$$

let c be the $(1-\alpha/2)$ quantile of t_{n-1}

$100 \cdot (1-\alpha)\%$ confidence interval

for $\mu_X - \mu_Y$ (population variances known)

$$\bar{x} - \bar{y} \pm c\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$$

$\Phi^{-1}(1-\alpha/2) = c$

TUTORIAL WEEK 5

Recap: Intervals (modules 3 & 4)

$100 \cdot (1-\alpha)\%$ confidence interval for σ^2

$$\left[\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a}\right]$$

where a and b are
the $\alpha/2$ and $1-\alpha/2$ quantiles of χ^2_{n-1}.

95% prediction interval for X^*,

$$\bar{x} \pm c\, \sigma \sqrt{1+\frac{1}{n}}$$

where c is the $1-\alpha/2$ quantile of
the standard normal distribution (known variance or large n) or t with n-1 dof (otherwise).
If population variance unknown,
estimate by sample variance.

$100 \cdot (1-\alpha)\%$ confidence interval for σ_X^2/σ_Y^2 as

$$\left[c\frac{s_x^2}{s_y^2}, d\frac{s_x^2}{s_y^2}\right]$$

where c and d are
the $\alpha/2$ and $1-\alpha/2$ quantiles of $\mathrm{F}_{m-1,n-1}$.
(n=sample size X, m=sample size Y)

$100 \cdot (1-\alpha)\%$ confidence for p

$$\hat{p} \pm c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where c is the $1-\alpha/2$ quantile of
the standard normal distribution

Sample size for proportions

In order to have $\hat{p} \pm \epsilon$ for a given ϵ
$n = c^2/(4\epsilon^2)$ as a conservative choice

$100 \cdot (1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$

$$\bar{x} - \bar{y} \pm c\sqrt{\mathrm{Var}(\bar{X}-\bar{Y})}$$

To calculate/estimate $\sqrt{\mathrm{Var}(\bar{X}-\bar{Y})}$

	$\sigma_X^2 \neq \sigma_Y^2$	$\sigma_X^2 = \sigma_Y^2 = \sigma^2$	c is the $1-\alpha/2$ quantile of
known σ_X^2, σ_Y^2	$\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}$	$\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}}$	standard normal
unknown σ_X^2, σ_Y^2	$\sqrt{\frac{s_X^2}{n}+\frac{s_Y^2}{m}}$ $r = \frac{\left(\frac{s_X^2}{n}+\frac{s_Y^2}{m}\right)^2}{\frac{s_X^4}{n^2(n-1)}+\frac{s_Y^4}{m^2(m-1)}}$	$s_P\sqrt{\frac{1}{n}+\frac{1}{m}}$ where $s_P = \sqrt{\frac{(n-1)s_X^2+(m-1)s_Y^2}{n+m-2}}$ $r = n+m-2$	large n,m: standard normal small n,m: t_r

TUTORIAL WEEK 6

Recap: regression (module 5)

Option 1: $Y = \alpha + \beta x_i + \varepsilon$

Option 2: $Y = \alpha_0 + \beta(x_i - \bar{x}) + \varepsilon$

where $\varepsilon \sim \mathrm{N}(0, \sigma^2)$.

Intercept: $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$

$\hat{\alpha}_0 = \bar{Y}$

$$\hat{\sigma}^2 = \frac{1}{n-2} D^2$$

$$D^2 = \sum_{i=1}^n R_i^2$$

$$\mathrm{var}(\hat{\alpha}) = \left(\frac{1}{n} + \frac{\bar{x}^2}{K}\right)\sigma^2$$

Slope:

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

$$= \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

To construct CIs for $\mu(x)$ recall:

$$\frac{\hat{\mu}(x) - \mu(x)}{\hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{K}}} \sim t_{n-2}$$

$100 \cdot (1-\alpha)\%$ CI for β

$$\hat{\beta} \pm c\,\frac{\hat{\sigma}}{\sqrt{K}}$$

where c is the $1-\alpha/2$ quantile of t_{n-2}

$K = \sum_i (x_i - \bar{x})^2$

A $100 \cdot (1-\alpha)\%$ PI for Y^* is given by:

$$\hat{\mu}(x^*) \pm c\,\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{K}}$$

where c is the $1-\alpha/2$ quantile of t_{n-2}

TUTORIAL WEEK 7

Recap: Hypothesis testing (module 6)

***p*-value:** the probability of getting the test statistic observed, or something more 'extreme' (with reference to the alternative hypothesis), assuming null hypothesis correct

Critical region: the region/values of the test statistic for which H_0 is rejected

Summary of outcomes table

		Decision made based on test statistic (calculated from data)	
		Do not reject H_0	Reject H_0
True state of the world (usually unknown)	H_0 is true	Correct! $1-\alpha$	**Type I error** Pr(reject H_0 | H_0 true)= α
	H_0 is false	**Type II error** Pr(do not reject H_0 | H_0 false)= β	Correct! power= $1-\beta$

Binomial approximation to Normal

If X~Bin(n, p) then X*~N(np,np(1-p))

requires large n and np, np(1-p) ≥ 5

Continuity correction: Pr(X=x)=Pr(x-0.5<X*<x+0.5)

Test stat: single variance

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Difference in proportions

Hypothesis testing: test statistic

$$\hat{p}_1 = y_1/n_1,\ \hat{p}_2 = y_2/n_2,\ \hat{p} = (y_1+y_2)/(n_1+n_2)$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n2)}} \approx \mathrm{N}(0,1)$$

Confidence interval for difference

$$\hat{p}_1 - \hat{p}_2 \pm c\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

where c is the $1-\alpha/2$ quantile of the standard normal distribution

Test stat: two variances

$$F = \frac{S_X^2}{S_Y^2} \sim \mathrm{F}_{n-1,m-1}$$

Test stat: single mean

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Small sample size and unknown σ

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

Test stat: two means

$$\text{test stat} = \frac{\bar{x} - \bar{y}}{\sqrt{\mathrm{Var}(\bar{X} - \bar{Y})}}$$

To calculate/estimate $\sqrt{\mathrm{Var}(\bar{X} - \bar{Y})}$

	$\sigma_X^2 \neq \sigma_Y^2$	$\sigma_X^2 = \sigma_Y^2 = \sigma^2$	Test stat distribution
known σ_X^2, σ_Y^2	$\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$	$\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}$	standard normal
unknown σ_X^2, σ_Y^2	$\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$ $r = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{s_X^4}{n^2(n-1)} + \frac{s_Y^4}{m^2(m-1)}}$	$s_P\sqrt{\frac{1}{n} + \frac{1}{m}}$ where $s_P = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}}$ $r = n+m-2$	large n,m: standard normal small n,m: t_r

TUTORIAL WEEK 8

Sign test

Compute, Y, the number of positive numbers amongst $X_1 - m_0, \ldots, X_n - m_0$

Under H_0, we have $Y \sim \mathrm{Bi}(n, 0.5)$

Wilcoxon signed-rank test statistic

$W = \sum_{i=1}^{n} W_i$ where $W_i = \mathrm{sgn}(X_i - m_0) \cdot \mathrm{rank}(|X_i - m_0|)$

$$Z = \frac{W - 0}{\sqrt{n(n+1)(2n+1)/6}} \approx \mathrm{N}(0,1)$$

Chi-squared test

$$Q_{k-1} = \sum_{i=1}^{k} \frac{(Y_i - np_i)^2}{np_i} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \approx \chi^2_{k-1}$$

Rule of thumb: need to have all $E_i = np_i \geqslant 5$

Chi-squared: test of independence

$$Q = \sum_i \sum_j \frac{(Y_{ij} - Y_{i\cdot} Y_{\cdot j}/n)^2}{Y_{i\cdot} Y_{\cdot j}/n} \approx \chi^2_{(r-1)(c-1)}$$

TUTORIAL WEEK 9

Recap: ANOVA (module 8)

Population	Sample	Statistics	
$N(\mu_1, \sigma^2)$	$X_{11}, X_{12}, \ldots, X_{1n_1}$	$\bar{X}_{1\cdot}$	S_1^2
$N(\mu_2, \sigma^2)$	$X_{21}, X_{22}, \ldots, X_{2n_2}$	$\bar{X}_{2\cdot}$	S_2^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$N(\mu_k, \sigma^2)$	$X_{k1}, X_{k2}, \ldots, X_{kn_k}$	$\bar{X}_{k\cdot}$	S_k^2
	Overall	$\bar{X}_{\cdot\cdot}$	

$n = n_1 + \cdots + n_k$ (total sample size)

$$\bar{X}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \quad \text{(group means)}$$

$$\bar{X}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij} = \frac{1}{n}\sum_{i=1}^{k} n_i \bar{X}_{i\cdot} \quad \text{(grand mean)}$$

Sum of squares

- The total SS is:

$$SS(TO) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{\cdot\cdot})^2$$

- The treatment SS, or between groups SS, is:

$$SS(T) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2 = \sum_{i=1}^{k} n_i(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$$

- The error SS, or within groups SS, is:

$$SS(E) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2 = \sum_{i=1}^{k}(n_i - 1)S_i^2$$

Hypothesis testing: means

$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots$ $\quad H_1$: At least one mean different

$$F = \frac{SS(T)/(k-1)}{SS(E)/(n-k)}$$

Under H_0, we have $F \sim F_{k-1,n-k}$

reject H_0 if $F > c$

Likelihood ratio tests

$H_0: \theta \in A_0$ versus $H_1: \theta \in A_1$

$$\lambda = \frac{L_0}{L_1} = \frac{\max_{\theta \in A_0} L(\theta)}{\max_{\theta \in A_1} L(\theta)}$$

- L is the likelihood function
- critical region of the form

$$\lambda \leqslant k$$

Choose k to give the desired significance level

Test statistics: regression

$$T_\alpha = \frac{\hat{\alpha} - \alpha}{\hat{\sigma}/\sqrt{n}} \sim t_{n-2}$$

$$T_\beta = \frac{\hat{\beta} - \beta}{\hat{\sigma}/\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \sim t_{n-2}$$

TUTORIAL WEEK 10

Recap: order statistics (module 9)

Sample from a continuous distribution with cdf $F(x)$ and pdf $f(x) = F'(x)$.

The cdf of $X_{(k)}$ is,

$$\begin{aligned} G_k(x) &= \Pr(X_{(k)} \leqslant x) \\ &= \sum_{i=k}^{n} \binom{n}{i} F(x)^i \left(1 - F(x)\right)^{n-i} \end{aligned}$$

$$g_k(x) = k \binom{n}{k} F(x)^{k-1} \left(1 - F(x)\right)^{n-k} f(x)$$

Special cases: minimum and maximum,

$$\begin{aligned} g_1(x) &= n \left(1 - F(x)\right)^{n-1} f(x) \\ g_n(x) &= n\, F(x)^{n-1} f(x) \\ \Pr(X_{(1)} > x) &= \left(1 - F(x)\right)^n \\ \Pr(X_{(n)} \leqslant x) &= F(x)^n \end{aligned}$$

Asymptotic distribution

For large sample sizes, it can be shown that

$$\hat{\pi}_p \approx \mathrm{N}\left(\pi_p, \frac{p(1-p)}{n f(\pi_p)^2}\right)$$

where f is the pdf of the population distribution

The median, $\hat{M} = \hat{\pi}_{0.5}$, is convenient special case,

$$\hat{M} \approx \mathrm{N}\left(m, \frac{1}{4n f(m)^2}\right)$$

CI for the median

In general, want i and j so that, to the closest possible extent,

$$\begin{aligned} \Pr(X_{(i)} < m < X_{(j)}) &= \Pr(i \leqslant W \leqslant j-1) \\ &= \sum_{k=i}^{j-1} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} \approx 1 - \alpha \end{aligned}$$

TUTORIAL WEEK 11

Recap: Bayesian Methods (module 10)

IDEA: Let's assume parameters themselves are also random variables (rather than unknown constants). So, both data/observations AND parameters have randomness.

Prior distribution: The distribution of the parameter BEFORE data is sampled. We may have some idea of the values that our parameter can take before collecting data, based on, e.g. past research
Posterior distribution: The distribution of the parameter AFTER data is sampled. This incorporates both prior beliefs and the information provided by the data.

Deriving the posterior distribution

Step 1: Identify the prior pdf/pmf $f(\theta)$
Step 2: Identify the likelihood of the data (the data might be a single observation or multiple observations) $f(y \mid \theta)$

For multiple observations: $f(y \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta)$

Step 3: Derive the posterior using the fact that:

$$f(\theta \mid y) \propto f(y \mid \theta) f(\theta)$$ ($\propto$ means is 'proportional to')

Write down and try to rearrange the right hand side until it looks like a familiar distribution. If this doesn't work, integrate/sum it with respect to theta over its range and equate to 1 to find the 'normalising constant' (your 'constant' may depend on the data).

Beta-binomial

prior, $\theta \sim \text{Beta}(\alpha, \beta)$

$X \sim \text{Bi}(n, \theta)$

posterior $\theta \mid X = x \sim \text{Beta}(x + \alpha, n - x + \beta)$

Pseudodata

Can think of the prior as being equivalent to unobserved data

A $\text{Beta}(\alpha, \beta)$ prior is equivalent to a sample of size of $\alpha + \beta$

Normal prior and data

Random sample: $X_1, \dots, X_n \sim \text{N}(\theta, \sigma^2)$, with σ^2 known

For simplicity, summarise the data by: $Y = \bar{X} \sim \text{N}(\theta, \sigma^2/n)$

Prior: $\theta \sim \text{N}(\mu_0, \sigma_0^2)$

Posterior: $\theta \mid y \sim \text{N}(\mu_1, \sigma_1^2)$

$$\mu_1 = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{y}{\sigma^2/n}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}} \quad \text{and} \quad \frac{1}{\sigma_1^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}$$

TUTORIAL WEEK 12

Asymptotic distribution of the MLE

$$\hat{\theta} \approx \mathrm{N}\left(\theta, \frac{1}{I(\theta)}\right) \quad \text{as } n \to \infty$$

where

$$I(\theta) = \mathbb{E}(V(\theta))$$

$$V(\theta) = -\frac{\partial U}{\partial \theta} = -\frac{\partial^2 \ell}{\partial \theta^2}$$

$$U(\theta) = \frac{\partial \ell}{\partial \theta} \qquad \text{(score function)}$$

Cramer-Rao lower bound

if we take any unbiased estimator T, then

$$\mathrm{var}(T) \geqslant \frac{1}{I(\theta)}$$

Factorisation theorem and sufficient statistics

Let $X_1, \ldots, X_n$ have joint pdf or pmf $f(x_1, \ldots, x_n \mid \theta)$

$Y = g(x_1, \ldots, x_n)$ is sufficient for θ if and only if

$$f(x_1, \ldots, x_n \mid \theta) = \phi\{g(x_1, \ldots, x_n) \mid \theta\}\, h(x_1, \ldots, x_n)$$

ϕ depends on $x_1, \cdots, x_n$ only through $g(x_1, \cdots, x_n)$ and h doesn't depend on θ.

Definition: the statistic $T = g(X_1, \ldots, X_n)$ is sufficient for an underlying parameter θ if the conditional probability distribution of the data $(X_1, \ldots, X_n)$, given the statistic $u(X_1, \ldots, X_n)$, does not depend on the parameter θ.

Sometimes need more than one statistic, e.g. T_1 and T_2, in which case we say they are jointly sufficient for θ