Cheat Sheet STAT111

Statistical Methods – Tommy Odland – Last edit: June 13, 2018

▷Basics

- The **conditional distribution** of Y given X is

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}.$$

  - It can be shown that $\mathrm{E}[Y] = \mathop{\mathrm{E}}_x \left[\mathop{\mathrm{E}}_y [Y|X] \right]$.
- The marginal distribution is $f_X(x) = \int f(x,y)\, dy$.
- The **variance** of a sum is given by

$$\mathrm{V}\left[\sum_i a_i X_i\right] = \sum_i \sum_j a_i\, a_j\, \mathrm{Cov}[X_i, X_j].$$

  - The **correlation** ρ_{XY} is a scaled covariance:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, \text{ where } \sigma_{XY} = \mathrm{Cov}[X_i, X_j].$$

- The expected value is $\mathrm{E}[h(x)] = \int p(x)h(x)\, dx$.
- A useful shortcut is $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$.
- When the **skew** $\mathrm{E}[(X-\mu)^3]/\sigma^3$ is positive, the right tail is longer or fatter.

▷Probability functions

- If a cont. mapping $Y = g(X)$ is monotonic with inverse $X = h(Y)$, then $f_Y(y) = f_X(x)\, |h'(y)|$.
  - The generalization uses the Jacobian $|\det \frac{\partial(\boldsymbol{x})}{\partial(\boldsymbol{y})}|$.
- The r'th order moment of X is $\mathrm{E}[X^r]$.
- The **moment generating function** (MGF) of X is defined as $M_X(t) = \mathrm{E}\left[e^{tX}\right]$.
  - It uniquely defines $f(x)$ if it's valid for a range including zero.
  - Defines the moments by $\partial_t^r M_X(t)|_{t=0} = \mathrm{E}[X^r]$.
  - If $R_X(t) = \ln M_X(t)$, then $\mu = R'_X(0)$ and $\sigma = R''_X(0)$.
  - $Y = \sum_i a_i X_i$, iff $M_Y(t) = \Pi_i M_{X_i}(a_i t)$
- **Discrete probability mass functions**
  - $\exp(\lambda) = \lambda e^{-\lambda x}$, $x \geq 0$
  - $\text{poisson}(\lambda) = \lambda^x e^{-\lambda}/x!$, $x \geq 0$
- **Continuous probability density functions**
  - The gamma function is $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x}\, dx$.
  - $\text{norm}(\mu, \sigma) = \exp(-(x-\mu)^2/2\sigma^2)/\sqrt{2\pi}\sigma$
  - X is lognormally distributed if $\log(X) \sim$ norm.
  - The Weibull distribution is defined for $x \geq 0$ as

$$f(x;\alpha,\beta) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \exp\left(-(x/\beta)^\alpha\right),$$

    where α is a location parameter, β is a scale parameter and $f(x;\alpha,\beta)$ may be shifted too.
  - The student-t distribution is the distribution of

$$T = \frac{Z}{\sqrt{X/\nu}} \cong \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

    where $Z \sim$ norm and $X \sim \chi^2_\nu$.
  - The Chi-squared distribution is the distribution of a sum of standard normals, i.e. $\chi^2_\nu = Z_1^2 + \cdots + Z_\nu^2$.
  - The f-squared distribution is the distribution of

$$F = \frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}.$$

  - The standard beta distribution on $[0,1]$ is

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}\,(1-x)^{\beta-1}$$

▷Order statistics

- The order statistics of $X_1, \ldots, X_n$ is obtained by sorting the random variables, yielding $Y_1, \ldots, Y_n$.
- The CDF of Y_1 (smallest value) is

$$P(Y_1 < y) = 1 - [1 - P(Y < y)]^n$$

- The CDF of Y_n (largest value) is

$$P(Y_n < y) = [P(Y \leq y)]^n$$

- The joint pdf of $Y_1, \ldots, Y_n$ is given by

$$g(y_1, \ldots, y_n) = n!\, f(y_1) \times \ldots \times f(y_n).$$

▷Point estimation and CIs

- A point estimate $\hat{\theta}$ is a sample based value for θ.
- **Moment estimators** equate theoretical moments $\mathrm{E}[X^r] = \int f(x) x^r\, dx$ with sample moments $n^{-1} \sum_i X_i^r$ to estimate θ. Simple, often biased.
- Three criteria for choosing $\hat{\theta}$:
  - No bias, choose $\mathrm{E}[\hat{\theta}] = \theta$.
  - If no bias, choose minimum variance $\mathrm{V}[\hat{\theta}]$.
  - Trade-off between bias and variance: $\min_{\hat{\theta}} \mathrm{MSE}$.
- **The mean squared error** of an estimator is

$$\mathrm{MSE}[\hat{\theta}] = \mathrm{E}[(\hat{\theta}-\theta)^2] = \mathrm{V}[\hat{\theta}] + \mathrm{E}[\hat{\theta}-\theta]^2 = \text{var} + \text{bias}^2$$

- Confidence interval for μ when σ is known:

$$\mu_0 \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Confidence interval for μ when σ is unknown:
$$\mu_0 \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$
- Confidence interval for σ^2:
$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$$
- Two approaches to **bootstrap** CI for $\bar{x}$:
  - Use $\bar{x} \pm z_{\alpha/2} s_\text{b}$, where $s_\text{b}^2 = \sum_i^n (\bar{x}_i^* - \bar{x}^*)^2/(n-1)$
  - Bootstrap, compute n $\bar{x}_i^*$'s, use percentiles.

▷Tests

- In a ***z*-test for differences**, the hypothesis is that $\mu_1 - \mu_2 = \Delta_0$, and σ_1 and σ_2 are both known. Combine standard deviations and test with z-test. Either X and Y are normal, or CLT applies.
- In a ***t*-test for differences**, the hypothesis is that $\mu_1 - \mu_2 = \Delta_0$, and $\sigma_{1,2}$ are both unknown. Combine s_1 and s_2 and use t-test. Degrees of freedom is found in an involved expression.
  - If $\sigma_1 = \sigma_2$, but both are unknown, pool them
  $$S_p^2 = \frac{m-1}{m+n-2}s_1^2 + \frac{n-1}{m+n-2}s_2^2,$$
  and employ a t-test with $m+n-2$ df.
- A **Type I** error is rejecting H_0 when H_0 is true. A **Type II** error is keeping H_0 when H_0 is false.
- The power function is $\beta(\mu) = P_\mu(\text{rejecting } H_0)$.

▷ANOVA

- In one-way **ANOVA**, the hypotheses are
$$H_0: \quad \mu_1 = \mu_2 = \cdots = \mu_I$$
$$H_0: \quad \text{at least 1 } \mu_i \neq \mu_j$$
Assumes normal distributions and same variance.
- The sum of squares are
$$\text{SSTr} = J\sum_i \left(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}\right)^2 \qquad \text{(between-group)}$$
$$\text{SSE} = \sum_i\sum_j \left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2 \qquad \text{(within-group)}$$
$$\text{SST} = \text{SSTr} + \text{SSE} \qquad \text{(total)}$$
$$\text{MSTr} = \frac{\text{SSTr}}{I-1} \quad \text{MSE} = \frac{\text{SSE}}{I(J-1)} \quad f = \frac{\text{MSTr}}{\text{MSE}}$$
  - The ratio is distributed as $f \sim F_{I-1,I(J-1)}$
  - Under the null hypothesis, SSTr $\sim \chi^2_{I-1}$, SSE $\sim \chi^2_{I(J-1)}$ and SST $\sim \chi^2_{IJ-1}$.
- To test for differences, sort μ_i's, form the interval
$$\bar{x}_{i\cdot} - \bar{x}_{j\cdot} \pm Q_{\alpha,I,I(J-1)}\sqrt{\text{MSE}/J}$$
and extend lines to test for significant differences.

▷Regression

- Assume $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, where ϵ_i are i.i.d. standard normals.
  - Implies $\text{E}[\epsilon_i] = 0$, $\text{V}[\epsilon] = \sigma^2$ and independence.
- $S_{xx} = \sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - n^{-1}\left(\sum_i x_i\right)^2$
- In $y = \hat{\beta}_0 + \hat{\beta}_1 x$, we compute $\hat{\beta}_1$ and $\hat{\beta}_0$ as:
  - $\hat{\beta}_1 = S_{xy}/S_{xx}$ and $\hat{\beta}_0 = \bar{y} - \bar{x}\beta_1$
- The coefficient of determination r^2 is given by
$$r^2 = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} = 1 - \frac{\text{res. } \sigma^2}{\text{tot. } \sigma^2}$$
The empirical correlation is $\rho = \sqrt{r^2}$.
- Interference about $\hat{\beta}_1$, i.e. $\hat{\beta}_1 \pm t_{\alpha/2,n-2}\, s_{\hat{\beta}_1}$.
  - $\text{V}[\hat{\beta}_1] = \sigma^2_{\hat{\beta}_1} = \sigma^2/S_{xx}$, where $\sigma^2 = \text{SSE}/(n-2)$.
- Sample correlation coefficient: $S_{xy}/(\sqrt{S_{xx}}\sqrt{S_{yy}})$.
- **Confidence interval** - interference of $\mu_{Y|X}$, which is deterministic. **Prediction interval** - interference of future Y values, which is stochastic.

▷Goodness-of-fit tests

- **Goodness-of-fit** tests are based on
$$\chi^2 = \sum_i^n \frac{(\text{obs}_i - \text{expected}_i)^2}{\text{expected}_i} \quad (\text{valid if } np_i \geq 5)$$
which has $\sim \chi^2_{n-1-m}$ degrees of freedom (m is the number of estimated parameters used to compute p_i), or alternatively $\sim \chi^2_{(I-1)(J-1)}$ if it's a contingency table.

▷Nonparametrics

- The **Wilcoxon signed rank test** has $H_0: \mu = \mu_0$. Subtract μ_0, rank the data by absolute value, then re-assign the signs. If H_0 is true, the sum of the positive ranks S_+ should be relatively small. Assumes symmetric distribution.
- The **Wilcoxon ranked sum test** has $H_0: \mu_A = \mu_B$. Assume m data points from A and n from B. Concatenate data, rank it, and sum ranks W associated with data from A. If H_0 is true, W will be somewhere in the middle of
$$\frac{m(m+1)}{2} \leq W \leq \frac{n(n+2m+1)}{2}$$
Assumes identical distribution shapes.

▷References

- Jay L. Devore. *Modern Mathematical Statistics with Applications.* 2nd ed. Springer Texts in Statistics. Dordrecht: Springer, 2011.