

## 1 Basic Properties

1.  $E(X) = \sum xp(x)$
2.  $Var(X) = \sum (x - \mu)^2 f(x)$
3. X is around  $E(X)$ , give or take  $SD(X)$
4.  $E(aX + bY) = aE(X) + bE(Y)$
5.  $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$
6.  $Var(X) = E(X^2) - [E(X)]^2$
7.  $Cov(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$
8.  $P(AB) = P(A)P(B)$  if A and B independent
9. RV is centered when  $E(X) = 0$ , and any RV can be centered via  $Y = X - E(X)$ , with SD and variance unaffected
10. In  $X = \mu + \epsilon$ ,  $\mu$  is the unknown constant of interest, and  $\epsilon$  represents random measurement error.
11. if X, Y are independent:
  - (a)  $M_{X+Y}(t) = M_X(t)M_Y(t)$
  - (b)  $E(XY) = E(X)E(Y)$ , converse is true if X and Y are bivariate normal, extends to multivariate normal

## 2 Approximations

### 2.1 Law of Large Numbers

Let  $X_1, X_2, \dots, X_n$  be IID, with expectation  $\mu$  and variance  $\sigma^2$ .  $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n]{\infty} \mu$ . Let  $x_1, x_2, \dots, x_n$  be realisations of the random variable  $X_1, X_2, \dots, X_n$ , then  $\overline{x}_n = \frac{1}{n} \sum_{i=1}^n x_n \xrightarrow[n]{\infty} \mu$

### 2.2 Central Limit Theorem

Let  $S_n = \sum_{i=1}^n X_i$  where  $X_1, X_2, \dots, X_n$  IID.  $\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow[n]{\infty} \mathcal{N}(0, 1)$

## 3 Distributions

### 3.1 Poisson( $\lambda$ )

$Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, \dots$   
 $E(X) = Var(X) = \lambda$

### 3.2 Normal $X \sim \mathcal{N}(\mu, \sigma^2)$

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), -\infty < x < \infty$

1. When  $\mu = 0$ ,  $f(x)$  is an even function, and  $E(X^k) = 0$  where  $k$  is odd
2.  $Y = \frac{X - E(X)}{SD(X)}$  is the standard normal

### 3.3 Gamma $\Gamma$

$g(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}, t \geq 0$   
 $\mu_1 = \frac{\alpha}{\lambda}, \mu_2 = \frac{\alpha(\alpha+1)}{\lambda^2}$

### 3.4 $\chi^2$ Distribution

Let  $\mathcal{Z} \sim \mathcal{N}(0, 1)$ ,  $\mathcal{U} = \mathcal{Z}^2$  has a  $\chi^2$  distribution with 1 d.f.  
 $f_{\mathcal{U}}(u) = \frac{1}{\sqrt{2\pi}} u^{-\frac{1}{2}} e^{-\frac{u}{2}}, u \geq 0$   
 $\chi_1^2 \sim \Gamma(\alpha = \frac{1}{2}, \lambda = \frac{1}{2})$   
Let  $U_1, U_2, \dots, U_n$  be  $\chi_1^2$  IID, then  $V = \sum_{i=1}^n U_i$  is  $\chi_n^2$  with n degree freedom,  $V \sim \Gamma(\alpha = \frac{n}{2}, \lambda = \frac{1}{2})$   
 $E(\chi_n^2) = n, Var(\chi_n^2) = 2n$   
 $M(t) = (1 - 2t)^{-\frac{n}{2}}$

### 3.5 t-distribution

Let  $\mathcal{Z} \sim \mathcal{N}(0, 1)$ ,  $\mathcal{U}_n \sim \chi_n^2$  be independent,  $t_n = \frac{\mathcal{Z}}{\sqrt{\mathcal{U}_n/n}}$  has a t-distribution with n d.f.

$f(t) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$

1. t is symmetric about 0
2.  $t_n \xrightarrow[n]{\infty} \mathcal{Z}$

### 3.6 F-distribution

Let  $U \sim \chi_m^2, V \sim \chi_n^2$  be independent,  $W = \frac{U/m}{V/n}$  has an F distribution with (m,n) d.f.

If  $X \sim t_n, X^2 = \frac{\mathcal{Z}/1}{\mathcal{U}_n/n}$  is an F distribution with (1,n) d.f, with  $w \geq 0$ :

$f(w) = \frac{\Gamma((n+1)/2)}{\Gamma(m/2)\Gamma(n/2)} \frac{m}{n} \frac{m}{2} w^{\frac{m}{2}-1} \left(1 + \frac{m}{n}w\right)^{-\frac{m+n}{2}}$   
For  $n > 2, E(W) = \frac{n}{n-2}$

## 4 Sampling

Let  $X_1, X_2, \dots, X_n$  be IID  $\mathcal{N}(\mu, \sigma^2)$ .

sample mean,  $\overline{X} = \frac{1}{n} \sum_{i=1}^n X_i$

sample variance,  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2$

### 4.1 Properties of $\overline{X}$ and $S^2$

1.  $\overline{X}$  and  $S^2$  are independent
2.  $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
3.  $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$
4.  $\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

## 4.2 Survey Sampling

In population of size  $N$ , we are interested in a variable  $x$ . The  $i$ th individual has fixed value  $x_i$ .

mean of population  $= \mu = \frac{1}{N} \sum_{i=1}^N x_i$

total of population  $= \tau = \sum_{i=1}^N x_i = \mu N$

SD of population  $= \sigma$

$\sigma^2 = \sum_{i=1}^N (x_i - \mu)^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \mu^2$

### 4.2.1 Dichotomous case

Population are members with value 0 or 1. Let  $p$  be the proportion of members with value 1.  $\mu = p, \sigma^2 = p(1-p)$

### 4.3 Simple Random Sampling (SRS)

Assume  $n$  random draws are made without replacement. (Not SRS, will be corrected for later).

#### 4.3.1 Lemma A

The draws  $X_i$  have the same distribution, and denote  $\xi_1, \xi_2, \dots, \xi_n$  as values assumed by the population, and let the number of members with value  $\xi_j$  be  $n_j$   
 $P(X_i = \xi_j) = \frac{n_j}{N}$   
 $E(X_i) = \mu, Var(x_i) = \sigma^2$

#### 4.3.2 Lemma B

For  $i \neq j, Cov(X_i, X_j) = -\frac{\sigma^2}{N-1}$

We use sample mean  $\overline{X}$  to estimate  $\mu$ :

$E(\overline{X}) = \mu$  from Lemma A, and

$Var(\overline{X}) = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1}\right)$  from Lemma B, where  $\frac{N-n}{N-1}$  is the finite population correction factor.

In 0-1 population, let  $\hat{p}$  be proportion of 1s in the sample:

$E(\hat{p}) = p, SD(\hat{p}) = \sqrt{\frac{p(1-p)}{n} \frac{N-n}{N-1}}$

#### 4.3.3 Estimation Problem

Let  $X_1, X_2, \dots, X_n$  be random draws with replacement. Then  $\overline{X}$  is an estimator of  $\mu$ . and the observed value of  $\overline{X}$ ,  $\bar{x}$  is an estimate of  $\mu$ .

#### 4.3.4 Standard Error (SE)

Since  $E(\overline{X}) = \mu$ , the estimator is unbiased.

The error in a particular estimate  $\overline{X}$  is unknown, but on average its size is about  $SD(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

Standard error of an  $\overline{X}$  is defined to be  $SD(\overline{X})$   
An unbiased estimator for  $\sigma^2$  is  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2$

param	est	SE	Est. SE
$\mu$	$\overline{X}$	$\frac{\sigma}{\sqrt{n}}$	$\frac{s}{\sqrt{n}}$
$p$	$\hat{p}$	$\sqrt{\frac{p(1-p)}{n}}$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}$

### 4.3.5 Without Replacement

SE is multiplied by  $\frac{N-n}{N-1}$ , because  $s^2$  is biased for  $\sigma^2$ :  $E(\frac{N-1}{N}s^2) = \sigma^2$ , but N is normally large.

### 4.3.6 Confidence Interval

An approximate  $1 - \alpha$  CI for  $\mu$  is  $(\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}})$

## 4.4 Measurement Error

Let  $x_1, x_2, \dots, x_n$  be independent measurements of unknown constant  $\mu$ .  $X_i = \mu + \epsilon_i$ . The errors are IID with expectation 0, and variance  $\sigma^2$ .  $x_i = \mu + e_i$ , where  $x_i$  and  $e_i$  are realisations of the RV. Then  $\bar{x}$  is an estimate of  $\mu$ , with SE  $\frac{\sigma}{\sqrt{n}}$ .

### 4.4.1 Biased Measurements

Let  $X = \mu + \epsilon$ , where  $E(\epsilon) = 0, Var(\epsilon) = \sigma^2$   
Suppose X is used to measure an unknown constant a,  $a \neq \mu$ .  $X = a + (\mu - a) + \epsilon$ , where  $\mu - a$  is the bias.

Mean square error (MSE) is  $E((X - a)^2) = \sigma^2 + (\mu - a)^2$

with n IID measurements,  $\bar{x} = \mu + \bar{\epsilon}$

$E((x - a)^2) = \frac{\sigma^2}{n} + (\mu - a)^2$

MSE = SE<sup>2</sup> + bias<sup>2</sup>, hence  $\sqrt{\text{MSE}}$  is a good measure of the accuracy of the estimate  $\bar{x}$  of a.

### 4.5 Estimation of a Ratio

Consider a population of  $N$  members, and two characteristics are recorded:  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n), r = \frac{\mu_y}{\mu_x}$ .

An obvious estimator of r is  $R = \frac{\bar{Y}}{\bar{X}}$

$Cov(\overline{X}, \overline{Y}) = \frac{\sigma_{xy}}{n}$ , where

$\sigma_{xy} := \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)(x_i - \mu_y)$  is the population covariance.

### 4.5.1 Properties

With SRS, the approx variance of  $R = \bar{Y}/\bar{X}$  is

$$\begin{aligned} Var(R) &\approx \frac{1}{\mu_x^2} (r^2 \sigma_X^2 + \sigma_Y^2 - 2r\sigma_{XY}) \\ &= \frac{1}{n} \frac{N-n}{N-1} \frac{1}{\mu_x^2} (r^2 \sigma_X^2 + \sigma_Y^2 - 2r\sigma_{XY}) \end{aligned}$$

Population coefficient  $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$

$$E(R) \approx r + \frac{1}{n} \left( \frac{N-n}{N-1} \right) \frac{1}{\mu_x^2} (r\sigma_x^2 - \rho\sigma_x\sigma_y)$$

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

### 4.5.2 Ratio Estimates

$$\bar{Y}_R = \frac{\mu_x}{\bar{X}} \bar{Y} = \mu_x R$$

$$Var(\bar{Y}_R) \approx \frac{1}{n} \frac{N-n}{N-1} (r^2 \sigma_x^2 + \sigma_y^2 - 2r\rho\sigma_x\sigma_y)$$

$$E(\bar{Y}_R) - \mu_y \approx \frac{1}{n} \frac{N-n}{N-1} \frac{1}{\mu_x} (r\sigma_x^2 - \rho\sigma_x\sigma_y)$$

The bias is of order  $\frac{1}{n}$ , small compared to its standard error.

$\bar{Y}_R$  is better than  $\bar{Y}$ , having smaller variance, when  $\rho > \frac{1}{2} \left( \frac{C_x}{C_y} \right)$ , where  $C_i = \sigma_i/\mu_i$

Variance of  $\bar{Y}_R$  can be estimated by

$$s_{\bar{Y}_R}^2 = \frac{1}{n} \frac{N-n}{N-1} (R^2 s_x^2 + s_y^2 - 2R s_{xy})$$

An approximate  $1 - \alpha$  C.I. for  $\mu_y$  is  $\bar{Y}_R \pm z_{\alpha/2} s_{\bar{Y}_R}$

## 5 Estimation

Let  $X_1, X_2, \dots, X_n$  be IID random variables with density  $f(x|\theta)$ , where  $\theta \in \mathcal{R}^P$  is an unknown constant. Realisations  $x_1, x_2, \dots, x_n$  will be used to estimate  $\theta$ , the estimate a realisation of RV  $\hat{\theta}$ . The bias and SE are:

$$\text{bias} = E(\hat{\theta}) - \theta, SE = SD(\hat{\theta})$$

### 5.1 Moments

Let  $X_1, X_2, \dots, X_n$  be IID with the same distribution as  $X$ .

$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$  is an estimator of  $\mu_k$ , where  $\mu_k$  is the  $k$ th moment. An estimate is also denoted  $\hat{\mu}_k$ .

### 5.2 Method of Moments

To estimate  $\theta$ , express it as a function of moments  $g(\hat{\mu}_1, \hat{\mu}_2, \dots)$

The bias and SE in an estimate, still depends on the unknown value of the constant. Suppose 1.67 and 0.38 are estimates of  $\lambda$  and  $\alpha$ .

Data is generated from  $\Gamma(1.67, 0.38)$ , and the MOM estimators are written as  $\widehat{1.67}$  and  $\widehat{0.38}$ . Because the sample size is large,  $(\hat{\lambda} - \lambda, \hat{\alpha} - \alpha) \approx (\widehat{1.67} - 1.67, \widehat{0.38} - 0.38)$

**Monte Carlo** is used to generate many realisations of  $\widehat{1.67}$  via the  $\Gamma(1.67, 0.38)$  distribution. With 10,000 realisations,  $bias(1.67) = E_{1.67, 0.38}(\widehat{1.67} - 1.67) \approx 0.09$

$$SE(1.67) = SD_{1.67, 0.38}(\widehat{0.38}) \approx 0.35$$

and  $\lambda$  is estimated as  $1.58 \pm 0.35$

$\bar{X} \xrightarrow[n]{\infty} \alpha/\lambda, \hat{\sigma}^2 \xrightarrow[n]{\infty} \alpha/\lambda^2$ , MOM estimators are consistent (asymptotically unbiased).

$$\text{Poisson}(\lambda): \text{bias} = 0, SE \approx \sqrt{\frac{\bar{x}}{n}}$$

$$N(\mu, \sigma^2): \mu = \mu_1, \sigma^2 = \mu_2 - \mu_1^2$$

$$\Gamma(\lambda, \alpha): \hat{\lambda} = \frac{\hat{\mu}_1}{\hat{\mu}_2 - \hat{\mu}_1^2} = \frac{\bar{X}}{\bar{\sigma}^2}, \hat{\alpha} = \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2} = \frac{\bar{X}^2}{\bar{\sigma}^2}$$

### 5.3 Maximum Likelihood Estimator (MLE)

Let  $f(\cdot|\theta) : \theta \in \Theta$  be a (identifiable) parametric identity

Suppose  $X_1, X_2, \dots, X_n$  are IID with density  $f(\cdot|\theta)$ , where  $\theta_0 \in \Theta$  is an unknown constant, we want to estimate  $\theta_0$  using realisations  $x_1, x_2, \dots, x_n$ .

$Pr(X_1 = x_1, X_2 = x_2, \dots) = \prod_{i=1}^n f(x_i|\theta)$  for a discrete distribution.

$$\theta \rightarrow L(\theta) = \prod_{i=1}^n f(x_i|\theta)$$

The maximum likelihood (ML) estimate of  $\theta_0$  is the number that maximises the likelihood over  $\theta$ .

The estimate is a realisation of the ML estimator  $\hat{\theta}_0$ , which can also be found by maximising

$$L(\theta) = \prod_{i=1}^n f(X_i|\theta)$$

The bias and SE are:

$$\text{bias} = E_{\theta_0}(\hat{\theta}_0) - \theta_0, SE = SD(\hat{\theta}_0)$$

#### 5.3.1 Poisson Case

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!}$$

$$l(\lambda) = \sum_{i=1}^n x_i \log \lambda - n\lambda - \sum_{i=1}^n \log x_i!$$

ML estimate of  $\lambda_0$  is  $\bar{x}$ . ML estimator is  $\hat{\lambda}_0 = \bar{X}$

#### 5.3.2 Normal case

$$l(\mu, \sigma) = -n \log \sigma - \frac{n \log 2\pi}{2} - \frac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^2}$$

$$\frac{\partial l}{\partial \mu} = \frac{\sum (X_i - \mu)}{\sigma^2} \implies \hat{\mu} = \bar{x}$$

$$\begin{aligned} \frac{\partial l}{\partial \sigma} &= \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^3} - \frac{n}{\sigma} \\ \implies \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{aligned}$$

#### 5.3.3 Gamma case

$$l(\theta) = n\alpha \log \lambda + (\alpha - 1) \sum_{i=1}^n \log X_i - \lambda \sum_{i=1}^n X_i - n \log \Gamma(\alpha)$$

$$\frac{\partial l}{\partial \alpha} = n \log \alpha + \sum_{i=1}^n \log X_i - \sum_{i=1}^n X_i - \frac{n}{\Gamma(\alpha)} \Gamma'(\alpha)$$

$$\frac{\partial l}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^n X_i$$

$$\hat{\lambda} = \frac{\hat{\alpha}}{\hat{x}}$$

bias and SE are estimated through Monte Carlo and Bootstrap methods.

#### 5.3.4 Multinomial Case

$$f(x_1, \dots, x_r) = \binom{n}{x_1, x_2, \dots, x_r} \prod_{i=1}^n p_i^{x_i}$$

where  $X_i$  is the number of times the value occurs, and not the number of trials. and  $x_1, x_2, \dots, x_r$  are non-negative integers summing to  $n$ .  $\forall i$ :

$$E(X_i) = np_i, Var(X_i) = np_i(1 - p_i)$$

$$Cov(X_i, X_j) = -np_i p_j, \forall i \neq j$$

$$l(p) = + \sum_{i=1}^{r-1} x_i \log p_i + x_r \log(1 - p_1 - \dots - p_{r-1})$$

$$\frac{\partial l}{\partial p_i} = \frac{x_i}{p_i} - \frac{x_r}{p_r} = 0 \text{ assuming MLE exists}$$

$$\frac{x_i}{p_i} = \frac{x_r}{p_r} \implies \hat{p}_i = \frac{x_i}{c}, c = \frac{x_r}{p_r}$$

$$\sum_{i=1}^r \hat{p}_i = \sum_{i=1}^r \frac{x_i}{c} = 1$$

$$\implies c = \sum_{i=1}^r x_i = n \implies \hat{p}_i = \frac{\bar{x}_i}{n}$$

same as MOM estimator.

#### 5.3.5 MLE vs MOM

1. ML estimates have smaller SEs than MOM estimates
2. In some cases bias and SE have to be computed numerically via methods like Newton-Raphson, and requires bootstrap and Monte Carlo

#### 5.3.6 Hardy-Weinberg Equilibrium

Let a locus have two alleles A and a, where the proportion of  $a$  in the population is  $\theta$ .

Assuming, the population is large, and mating is random, then in the next generation, the proportion of  $a$  alleles is the sum of 2 Be RV,  $Bin(2, \theta)$  and the number of  $a$  alleles is  $Bin(2n, \theta)$

### 5.3.7 CIs in MLE

When sample size is large,  $\hat{\theta}_0$  is approximately normal.

$$\frac{\hat{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

Given the realisations  $\bar{x}$  and  $s$ ,

$$\left( \bar{x} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}} \right)$$

is the exact  $1 - \alpha$  CI for  $\mu$ .

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-1}$$

$$\left( \frac{n\hat{\sigma}^2}{\chi_{n-1, \alpha/2}^2}, \frac{n\hat{\sigma}^2}{\chi_{n-1, 1-\alpha/2}^2} \right)$$

is the exact  $1 - \alpha$  CI for  $\sigma$ .