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. $Var(aX+bY) = a^2Var(X) + b^2Var(Y)$
6. $Var(X) = E(X^2) - [E(X)]^2$
7. $Cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2)$
8. 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 $X_1, X_2, ..., X_n$ be IID, with expectation μ and variance σ^2. $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n]{\infty} \mu$. Let $x_1, x_2, ..., x_n$ be realisations of the random variable $X_1, X_2, ..., X_n$, 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 $X_1, X_2, ..., X_n$ IID. $\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow[n]{\infty} \mathcal{N}(0,1)$

3 Distributions

3.1 Poisson(λ)

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

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 Γ

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

$\mu_1 = \frac{\alpha}{\lambda}, \mu_2 = \frac{\alpha(\alpha+1)}{\lambda^2}$

3.4 χ^2 Distribution

Let $\mathcal{Z} \sim \mathcal{N}(0,1)$, $\mathcal{U} = \mathcal{Z}^2$ has a χ^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 $U_1, U_2, ..., U_n$ be χ_1^2 IID, then $V = \sum_{i=1}^n U_i$ is χ_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 $X_1, X_2, ..., X_n$ be IID $\mathcal{N}(\mu, \sigma^2)$.
sample mean, $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$
sample variance, $S^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$

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 Simple Random Sampling (SRS)

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

4.2.1 Summary of Lemmas

- $P(X_i = \xi_j) = \frac{n_j}{N}$: Lemma A
- For $i \neq j$, $Cov(X_i, X_j) = -\frac{\sigma^2}{N-1}$: Lemma B

4.2.2 Estimation Problem

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

4.2.3 Standard Error (SE)

SE of an $\overline{X}$ is defined to be $SD(\overline{X})$.

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

4.2.4 Without Replacement

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

4.2.5 Confidence Interval

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

4.3 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, $\overline{x} = \mu + \overline{\epsilon}$
$E((x-a)^2) = \frac{\sigma^2}{n} + (\mu - a)^2$
MSE $=$ SE^2 + bias^2, hence $\sqrt{\text{MSE}}$ is a good measure of the accuracy of the estimate $\overline{x}$ of a.

4.4 Estimation of a Ratio

Consider a population of N members, and two characteristics are recorded: $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$, $r = \frac{\mu_y}{\mu_x}$.
An obvious estimator of r is $R = \frac{\overline{Y}}{\overline{X}}$
$Cov(\overline{X}, \overline{Y}) = \frac{\sigma_{xy}}{n}$, where
$\sigma_{xy} := \frac{1}{N}\sum_{i=1}^N (x_i - \mu_x)(x_i - \mu_y)$ is the population covariance.

4.4.1 Properties

$Var(R) \approx \frac{1}{\mu_{\overline{x}}^2}\left(r^2\sigma_{\overline{X}}^2 + \sigma_{\overline{Y}}^2 - 2r\sigma_{\overline{XY}}\right)$
Population coefficient $\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$
$E(R) \approx r + \frac{1}{n}\left(\frac{N-n}{N-1}\right)\frac{1}{\mu_x^2}\left(r\sigma_x^2 - \rho\sigma_x\sigma_y\right)$
$s_{xy} = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$

4.4.2 Ratio Estimates

$\overline{Y}_R = \frac{\mu_x}{\overline{X}}\overline{Y} = \mu_x R$
$Var(\overline{Y}_R) \approx \frac{1}{n}\frac{N-n}{N-1}(r^2\sigma_x^2 + \sigma_y^2 - 2r\rho\sigma_x\sigma_y)$
$E(\overline{Y}_R) - \mu_y \approx \frac{1}{n}\frac{N-n}{N-1}\frac{1}{\mu_x}\left(r\sigma_x^2 - \rho\sigma_x\sigma_y\right)$
The bias is of order $\frac{1}{n}$, small compared to its standard error.
$\overline{Y}_R$ is better than $\overline{Y}$, having smaller variance, when $\rho > \frac{1}{2}\left(\frac{C_x}{C_y}\right)$, where $C_i = \sigma_i/\mu_i$
Variance of $\overline{Y}_R$ can be estimated by
$s_{\overline{Y}_R}^2 = \frac{1}{n}\frac{N-n}{N-1}\left(R^2s_x^2 + s_y^2 - 2Rs_{xy}\right)$
An approximate $1-\alpha$ C.I. for μ_y is $\overline{Y}_R \pm z_{\alpha/2}s_{\overline{Y}_R}$

5 Method of Moments

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

5.1 Monte Carlo

Monte Carlo is used to generate many realisations of random variable.
$\overline{X} \xrightarrow[n]{\infty} \alpha/\lambda, \hat{\sigma}^2 \xrightarrow[n]{\infty} \alpha/\lambda^2$, MOM estimators are consistent (asymptotically unbiased).
Poisson(λ): bias $= 0, SE \approx \sqrt{\frac{\overline{x}}{n}}$
$N(\mu, \sigma^2)$: $\mu = \mu_1$, $\sigma^2 = \mu_2 - \mu_1^2$
$\Gamma(\lambda, \alpha)$: $\hat{\lambda} = \frac{\hat{\mu}_1}{\hat{\mu}_2 - \hat{\mu}_1^2} = \frac{\overline{X}}{\hat{\sigma}^2}, \hat{\alpha} = \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2} = \frac{\overline{X}^2}{\hat{\sigma}^2}$

6 Maximum Likelihood Estimator (MLE)

6.1 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 λ_0 is $\overline{x}$. ML estimator is $\hat{\lambda}_0 = \overline{X}$

6.2 Normal case

$l(\mu, \sigma) = -n\log\sigma - \frac{n\log 2\pi}{2} - \frac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^2}$
$\frac{\partial l}{\partial\mu} = \frac{\sum(X_i - \mu)}{\sigma^2} \implies \hat{\mu} = \overline{x}$
$\frac{\partial l}{\partial\sigma} = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^3} - \frac{n}{\sigma}$
$\implies \hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$

6.3 Gamma case

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

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

6.4 Multinomial Case

$f(x_1,...,x_r) = \binom{n}{x_1,x_2,...x_r}\prod_{i=1}^n p_i^{X_i}$
where X_i is the number of times the value occurs, and not the number of trials. and $x_1,x_2,...x_r$ are non-negative integers summing to n. $\forall i$:
$E(X_i)=np_i, Var(X_i)=np_i(1-p_i)$
$Cov(X_i,X_j) = -np_ip_j, \forall i\neq j$
$l(p) = +\sum_{i=1}^{r-1} x_i\log p_i + x_r\log(1-p_1-...-p_{r-1})$
$\frac{\partial l}{\partial p_i} = \frac{x_i}{p_i} - \frac{x_r}{p_r} = 0$ assuming MLE exists
$\frac{x_i}{\hat p_i} = \frac{x_r}{\hat p_r} \implies \hat p_i = \frac{x_i}{c}, c = \frac{x_r}{\hat p_r}$
$\sum_{i=1}^r \hat p_i = \sum_{i=1}^r \frac{x_i}{c} = 1$
$\implies c = \sum_{i=1}^r x_i = n \implies \hat p_i = \frac{\overline{x}_i}{n}$
same as MOM estimator.

6.5 CIs in MLE

$\frac{\hat X - \mu}{s/\sqrt n} \sim t_{n-1}$
Given the realisations $\overline x$ and s, $\overline x \pm t_{n-1,\alpha/2}\frac{s}{\sqrt n}, \overline x + t_{n-1,\alpha/2}\frac{s}{\sqrt n}$ is the exact $1-\alpha$ CI for μ.
$\frac{n\hat\sigma^2}{\sigma^2} \sim \chi_{n-1}$, $\frac{n\hat\sigma^2}{\chi^2_{n-1,\alpha/2}}, \frac{n\hat\sigma^2}{\chi^2_{n-1,1-\alpha/2}}$ is the exact $1-\alpha$ CI for σ.

7 Fisher Information

$I(\theta) = -E\left(\frac{\partial}{\partial\theta^2}\log f(x|\theta)\right)$

Distribution	MLE	Variance
Po(λ)	X	λ
Be(p)	X	$p(1-p)$
Bin(n,p)	$\frac{X}{n}$	$\frac{p(1-p)}{n}$
HWE tri	$\frac{X_2+2X_3}{n}$	$\frac{\theta(1-\theta)}{n}$

General trinomial: $\left(\frac{X_1}{n}, \frac{X_2}{n}\right)$

$$\begin{bmatrix} p_1(1-p_1) & -p_1p_2 \\ -p_1p_2 & p_2(1-p_2)\end{bmatrix}\frac{1}{n}$$

In all the above cases, var($\hat\theta$) = $I(\theta)^{-1}$.

8 Asymptotic Normality of MLE

As $n\to\infty$, $\sqrt{nI(\theta)}(\hat\theta-\theta)\to N(0,1)$ in distribution, and hence $\hat\theta \sim N\left(\theta, \frac{I(\theta)^{-1}}{n}\right)$
As $\hat\theta \xrightarrow[n]{\infty} \theta$, MLE is consistent.
SE of an estimate of θ is the SD of the estimator $\hat\theta$, hence $SE = SD(\hat\theta) = \sqrt{\frac{I(\theta)^{-1}}{n}} \approx \sqrt{\frac{I(\hat\theta)^{-1}}{n}}$
$1-\alpha$ CI $\approx \hat\theta \pm z_{\alpha/2}\sqrt{\frac{I(\theta)^{-1}}{n}}$

9 Efficiency

Cramer-Rao Inequality: if θ is unbiased, then $\forall\theta\in\Theta$, $var(\hat\theta)\geq I(\hat\theta)^{-1}/n$, if = then $\hat\theta$ is efficient.
$eff(\hat\theta) = \frac{I(\hat\theta)^{-1}/n}{var(\hat\theta)} < 1$

10 Sufficiency

10.1 Characterisation

Let $S_t = x : T(x) = t$. The sample space of X, S is the disjoint union of S_t across all possible values of T.
T is sufficient for θ if $\exists q()$ s.t. $\forall x \in S_t, f_\theta(X|T=t) = q(x)$.

10.2 Factorisation Theorem

T is sufficient for θ iff $\exists g(t,\theta), h(x)$ s.t. $\forall\theta\in\Theta, f_\theta(x) = g(T(x),\theta)h(x)\forall x$

10.3 Rao-Blackwell Theorem

Let $\hat\theta$ be an estimator of θ with finite variance, T be sufficient for θ. Let $\tilde\theta = E[\hat\theta|T]$. Then for every $\theta\in\Theta$, $E\left(\hat\theta-\theta\right)^2 \leq E\left(\hat\theta-\theta\right)^2$. Equality holds iff $\hat\theta$ is a function of T.

10.4 Random Conditional Expectation

1. $E(X)=E(E(X|T))$
2. $var(X) = var(E(X|T)) + E(var(X|T))$
3. $var(Y|X) = E(Y^2|X) - E(Y|X)^2$
4. $E(Y)=Y, var(Y)=0$ iff Y is a constant

11 Hypothesis Testing

Let $X_1...X_n$ be IID with density $f(x|\theta)$. null $H_0:\theta=\theta_0$, $H-1:\theta=\theta_1$. Critical region is R_n. $size = P_0(X\in R)$ and $power = P_1(X\in R)$.
$\Lambda(x) = \frac{f_0(x_1)...f_0(x_n)}{f_1(x_1)...f_1(x_n)}$. Critical region $x:\Lambda(x)<c_\alpha$, and among all tests with this size, it has the maximum power (Neyman-Pearson Lemma).
A hypothesis is simple if it completely specifies the distibution of the data.
$H_1:\mu>\mu_0$: Critical region $\{\bar x > \mu_0 + z_\alpha\frac{\sigma}{\sqrt n}\}$, the power is a function of μ, and this is uniformly the most powerful test for size $\leq\alpha$.
$H_1:\mu\neq\mu_0$: Critical region $\{|\bar x-\mu_0|>c\}, c = z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt n}$, but not uniformly most powerful.
The $(1-\alpha)$ CI for μ consists of precisely the values μ_0 for which $H_0:\mu=\mu_0$ is not rejected against $H_1:\mu\neq\mu_0$. Exact for normal with known variance, approx. in others.

11.1 p-value

the probability under H_0 that the test statistic is more extreme than the realisation. (A, B): $p = p_0(\bar X > \bar x) = P(Z > \frac{\bar x-\mu_0}{\sigma/\sqrt n})$. (C): $p = P_0(|\bar X-\mu_0| > |\bar x-\mu_0|)$. The smaller the p-value, the more suspicious one should be about H_0. If size is smaller than p-value, do not reject H_0.

12 Generalized Likelihood Ratio

$\Lambda^* = \frac{\max_{\theta\in\omega_0}L(\theta)}{\max_{\theta\in\Omega}L(\theta)}$, $\Omega=\omega_0\cup\omega_1$. The closer Λ is to 0, the stronger the evidence for H_1.

12.1 Large-sample null distribution of Λ

Under H_0, when n is large, $-2\log\Lambda = \chi^2_k$, where $k = \dim(\Omega)-\dim(\omega_0)$.
Normal (C): $p = P\left(\chi^2_1 > \frac{(\bar x-\mu_0)^2}{\sigma^2/n}\right)$
Multinomial: $\Lambda = \prod_{i=1}^r\left(\frac{E_i}{X_i}\right)_i^X$ where $E_i = np_i(\hat\theta)$ is the expected frequency of the ith event under H_0. $-2\log\Lambda\approx\sum_{i=1}^r\frac{(X_i-E_i)^2}{E_i}$, which is the Pearson chi-square statistic, written as X^2.

12.2 Poisson Dispersion Test

For $i=1...n$ let $X_i\sim Poisson(\lambda_i)$ are independent.
$w_0 = \{\tilde\lambda|\lambda_1=\lambda_2=...=\lambda_n\}$
$w_1 = \{\tilde\lambda|\lambda_i\neq\lambda_j \text{ for some } i,j\}$
$-2\log\Lambda\approx\frac{\sum_{i=1}^n(X_i-\bar X)^2}{\bar X}$. For large n, the null distribution of $-2\log\Lambda$ is approximately χ^2_{n-1}

13 Comparing 2 samples

13.1 Normal Theory: Same Variance

$X_1,...,X_n$ be i.i.d $N(\mu_X,\sigma^2)$ and $Y_1,...,Y_m$ be i.i.d $N(\mu_Y,\sigma^2)$, independent. $H_0:\mu_X-\mu_Y=d$

13.1.1 Known Variance

$Z := \frac{\bar X-\bar Y-(\mu_X-\mu_Y)}{\sigma\sqrt{\frac1n+\frac1m}}$ and reject H_0 when $|Z| > z_{\alpha/2}$

13.1.2 Unknown Variance

$s_p^2 = \frac{(n-1)s_X^2+(m-1)s_Y^2}{m+n-2}$ where $\$\mathrm{s}_{\mathrm{X}}^2 = 1\frac{}{n-1\sum_{\mathrm{i}=1}^{\mathrm{n}}}(\mathrm{X_i}\text{-})^2$. s_p^2 is an unbiased estimator of σ^2. s_X within factor of 2 from s_Y.
$t := \frac{\bar X-\bar Y-(\mu_X-\mu_Y)}{s_p\sqrt{\frac1n+\frac1m}}$ follows a t distribution with $m+n-2$ d.f.
If two-sided: reject H_0 when $|t| > t_{n+m-2,\alpha/2}$.
If one-sided, e.g $H_1:\mu_X>\mu_Y$, reject H_0 when $t > t_{n+m-2,\alpha}$.

13.1.3 CI

$\frac{\bar X-\bar Y}{\pm}z_{\alpha/2}\cdot\sigma\sqrt{\frac1n+\frac1m}$ if σ is known, or $\frac{\bar X-\bar Y}{\pm}t_{m+n-2,\alpha/2}\cdot s_p\sqrt{\frac1n+\frac1m}$ if σ is unknown.

13.1.4 Unequal Variance

$Z := \frac{\bar X-\bar Y-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}$
$t := \frac{\bar X-\bar Y-(\mu_X-\mu_Y)}{\sqrt{\frac{s_X^2}{n}+\frac{s_Y^2}{m}}}$, with $df = \frac{(a+b)^2}{\frac{a^2}{n-1}+\frac{b^2}{m-1}}$ where $a = \frac{s_X^2}{n}$ and $b = \frac{s_Y^2}{m}$

13.2 Mann-Whitney Test

We take the smaller sample of size n_1, and sum the ranks in that sample. $R' = n_1(m+n+1)-R$, and $R* = min(R',R)$, we reject $H_0: F=G$ if $R*$ is too small.
Test works for all distributions, and is robust to outliers.

13.3 Paired Samples

(X_i,Y_i) are paired and related to the same individual. (X_i,Y_i) is independent from (X_j,Y_j). Compute $D_i = Y_i - X_i$, To test $H_0:\mu_D=d$, $t = \frac{\bar D-\mu_D}{s_D/\sqrt n}$.
$1-\alpha$ CI: $\bar D\pm t_{n-1,\alpha/2}S_D/\sqrt n$

13.4 Ranked Test

W_+ is the sum of ranks among all positive D_i and W_i is the sum of ranks among all negative D_i. We want to reject H_0 if $W = min(W_+,W_-)$ is too large.