Review: Statistics dual $\longleftrightarrow$ Probability

outcomes $\to$ distributions   distributions $\to$ outcomes

A & B events are independent if:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

X & Y RVs are indep if:

$$\underbrace{f_{X,Y}(x,y)}_{\text{joint density}} = \underbrace{f_X(x) \cdot f_Y(y)}_{\text{marginal densities}}, \quad f_X(x) = \int_Y f_{X,Y}(x,y)\,dy.$$

$$\text{Indep} \Rightarrow \mathrm{Cov}(X,Y) = 0 = \mathbb{E}((X-\mu)(Y-\nu)) \leftarrow \text{def.}$$

$$= \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

$$\mathrm{Var}(X) = \mathbb{E}((X-\mu)^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

$$M_X(t) = \text{moment generating function} = \mathbb{E}(e^{tX}) = \int e^{tx} f_X(x)\,dx.$$

$$e^{tX} = 1 + tX + \frac{t^2 X^2}{2} + \frac{t^3 X^3}{3!} + \cdots$$

$$= \cancel{\int f_X(x)\,dx}^{\,1} + t\int x f_X(x)\,dx + \frac{t^2}{2}\int x^2 f_X(x)\,dx + \cdots$$

$$\left.\frac{\partial^k}{\partial t^k} M_X(t)\right|_{t=0} = \mathbb{E}(X^k).$$

$$f_Y(y \mid X=x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$ conditional probability density.

$$\mathbb{E}(Y \mid X) = \int y\, f_Y(y \mid X=x)\,dy.$$ conditional expectation (RV of X).

Important Distributions: $\mathcal{N}(\mu, \sigma^2)$, Bernoulli θ,

Binomial $= \sum$ Bernoulli, exponential, Gamma, ...

(∞) population sample $X_1, \ldots, X_n$ iid RV.

$\overline{X}$ = sample mean $= \frac{1}{n}\sum X_i$. $\mathbb{E}(\overline{X}) = \mu$.

$$\mathrm{Var}(\overline{X}) = \mathbb{E}((\overline{X}-\mu)^2) = \mathbb{E}(\overline{X}^2) - \mu^2 = \frac{\sigma^2}{n}.$$

$$= \frac{1}{n^2}\sum\left(X_i^2 + 2X_iX_j\right). \qquad \mathrm{Cov}(X_i, X_j) = 0.$$

Fact: X & Y indep $\Rightarrow \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Chebyshev Thm: $\mathbb{P}(|X-\mu| > k\sigma) < \frac{1}{k^2}$.

pf:
$$\int_{|x-\mu|>k\sigma} f_X(x)\,dx \le \int_{|x-\mu|>k\sigma} \frac{(x-\mu)^2}{k^2\sigma^2} f_X(x)\,dx$$
$$\le \frac{1}{k^2\sigma^2}\int_{\Omega} (x-\mu)^2 f_X(x)\,dx = \frac{1}{k^2}$$

LLN: $\mathbb{P}(|\overline{X}-\mu| > c) \le \frac{c^2}{\sigma^2 n} \to 0$. $\left(\frac{k\sigma}{\sqrt{n}} = c\right)$.

CLT: $$\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} \xrightarrow{(d)} \mathcal{N}(0,1).$$

pf used characteristic function $\varphi_X(t) = \mathbb{E}(e^{itX})$.

Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. (Bessel correction).

$\mathbb{E}(S^2) = \sigma^2$

Specify to normal sample $X_i \sim N(\mu, \sigma^2)$.

$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$, $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{(n-1)}$ deg freedom.

$Z \sim N(0,1)$, $\mathbb{P}(Z^2 \le a) = \mathbb{P}(-\sqrt{a} \le Z \le \sqrt{a}) = 2\mathbb{P}(0 \le Z \le \sqrt{a})$

$$= 2\int_0^{\sqrt{a}} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx = \cancel{2}\int_0^a \frac{e^{-\frac{u}{2}}}{\sqrt{2\pi}} \cdot \frac{1}{\cancel{2}} u^{1/2} \frac{du}{u}$$

$u = x^2$, $x = u^{1/2}$, $dx = \frac{1}{2}u^{-1/2}du$ $\qquad$ " $f_{\chi^2_1}$ $(u>0)$.

If $Y = Z_1^2 + \dots + Z_n^2 \sim \chi^2_n$.

$$M_{\chi^2_1}(t) = \int_0^\infty \frac{e^{tu}}{\sqrt{2\pi}} e^{-\frac{u}{2}} u^{1/2} \frac{du}{u} \qquad (t < 1/2).$$

$V = (\frac{1}{2} - t)u$.

$$= \int_0^\infty \frac{e^{-v}}{\sqrt{2\pi}} \frac{v^{1/2}}{(\frac{1}{2}-t)^{1/2}} \frac{dv}{v} = \frac{(\frac{1}{2}-t)^{-1/2}}{\sqrt{2\cancel{\pi}}} \cancel{\Gamma(\tfrac{1}{2})}$$

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$$= (1 - 2t)^{-1/2}.$$

$$M_Y(t) = (1-2t)^{-n/2} \Rightarrow \text{pdf } f_Y(y) = \frac{e^{-\frac{y}{2}}\, y^{n/2}\, \frac{1}{y}}{2^{n/2}\,\Gamma\left(\frac{n}{2}\right)}.$$

$$\frac{(n-1)S^2}{\sigma^2} = \cancel{\frac{1}{n-1}} \sum_{i=1}^{n} \cancel{(X_i - \bar{X})}^2$$

$$\left(\frac{X_i-\mu}{\sigma} - \frac{\bar{X}-\mu}{\sigma}\right)^2 \sim \chi^2_{n-1}.$$

Want: $\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \rightsquigarrow \frac{\bar{X}-\mu}{\sqrt{S^2/n}}$ don't know σ^2, approx by S^2.

$\leftarrow$ no longer normal.

T_n-distribution: $T = Z \cdot \sqrt{\frac{n}{Y}}$

$Y \leftarrow \chi^2_n$

$Z \leftarrow N(0,1)$

$$\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sqrt{\frac{\sigma^2(n-1)}{S^2(n-1)}}$$

Fact: $\bar{X}$ & S^2 indep.

$$\text{pdf } f_{T_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{t^2}{n}\right)^{-\frac{(n+1)}{2}} \xrightarrow[n \text{ large}]{} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}.$$

Order statistics: $X_1, \dots, X_n$, $Y_1 \le \dots \le Y_n$

$$f_{Y_k}(y) = \underline{\binom{n}{k-1}} \left(\int_{-\infty}^{y} f_X(x)\,dx\right)^{k-1} \cdot \underline{(n-k+1)} \cdot f_X(y) \left(\int_{y}^{\infty} f_X(x)\,dx\right)^{n-k}.$$

Estimator $\hat{\theta}(X_1, \dots, X_n)$, unbiased if $\mathbb{E}(\hat{\theta}) = \theta$.

E.g.: Uniform on $[0, \theta)$, $\hat{\theta}_1 = 2\bar{X}$, $\hat{\theta}_2 = \frac{n+1}{n} Y_n$.

Relative efficiency of $\hat{\theta}_1$ to $\hat{\theta}_2$ is $\dfrac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)} > 1$.

Cramer Rao Ineq: $Var(\hat{\theta}) \geq \dfrac{1}{n \underbrace{\mathbb{E}\left(\left(\frac{\partial}{\partial \theta} \log f_X(x;\theta)\right)^2\right)}_{\text{information}}}$

unbiased estimator + (*)

optimal.

E.g.: $X_i \sim N(\mu, \sigma^2)$, $\bar{X} = \hat{\mu}$

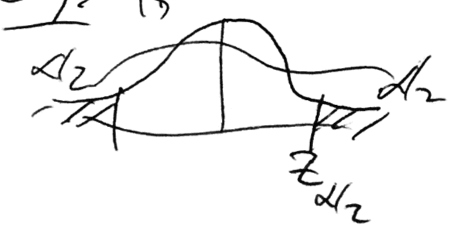
Ratio of variances $F = \dfrac{U/n}{V/m}$, $U \leftarrow \chi^2_n$, $V \leftarrow \chi^2_m$, indep.

Robust methods for ~~obtain~~ generating point estimators:

- moment method
- max likelihood method.

(Confidence) Interval estimation: I is a $(1-\alpha)100\%$ confidence interval for $\hat{\theta}$ if: $\mathbb{P}(\theta \in I(\hat{\theta})) = 1 - \alpha$.

E.g.: Normal $\mathbb{P}\left(\left|\dfrac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}\right| \leq z_{\alpha/2}\right) = 1 - \alpha$.



interval for μ, $\hat{\mu} = \bar{X}$, if σ is known.

If σ unknown, T-distr $\sigma^2 \rightsquigarrow S^2$.

For σ: $\frac{(n-1)}{\sigma^2} S^2 \sim \chi^2$.

Difference of μ's $\underline{\mu_X - \mu_Y}$ (if σ_X, σ_Y known, use Z; if not but $\sigma_X = \sigma_Y = \sigma$ use T, pooled estimates)

Ratio of σ's: F-distr.

<u>Proportions:</u>

$$\frac{n\bar{X} - n\theta}{\sqrt{n\theta(1-\theta)}} \approx N(0,1).$$

$\bar{X}$.

<u>Difference of Prop:</u> same gives Z-value

<u>Hypothesis testing:</u> H_0 = null hypothesis (want to reject).

H_1: alternative. C = critical region where H_0 rejected.

α = size of Type I error = $\mathbb{P}(C; H_0)$.

β = prob of Type II error, $1-\beta$ = power = $\mathbb{P}(C; H_1)$.

<u>Neyman Pearson Lemma:</u> $H_0: \theta = \theta_0$ $H_1: \theta = \theta_1$. If

$$\frac{\mathcal{L}_0}{\mathcal{L}_1} \leq k \ \forall \vec{x} \in C \ \& \ \frac{\mathcal{L}_0}{\mathcal{L}_1} \geq k \ \forall \vec{x} \in C^c$$

Then C is a most powerful region of its size.

For composite hypotheses, $H_0 : \theta \in \Theta_0$, $H_1 : \theta \in \Theta_1$.

Likelihood ratio test $C : \{ \lambda \le k \}$, $k < 1$, where

$$\lambda = \frac{\max \mathcal{L}_0}{\max \mathcal{L}}.$$

$\boxed{\text{Wilks's Thm: } -2\log\lambda \xrightarrow{(d)} \chi^2_1}$

<u>Def (p-values)</u>: least size α s.t. H_0 rejected.

Family of $C(\alpha)$, normal (σ known/unknown), diff of means (σ_x, σ_y known, unknown $\sigma_x = \sigma_y$ T-value), variances (χ^2), ratios of variances (F), proportions (Z), Multiple samples differences of proportions.

	success	failure	total
Sample 1	f_1	$n_1 - f_1$	n_1
Sample 2	f_2	$n_2 - f_2$	n_2
Sample 3	f_3	$n_3 - f_3$	n_3

$H_0 : \theta_1 = \theta_2 = \theta_3 = \theta_0$.

$$\hat{\Theta} = \frac{f_1 + f_2 + f_3}{n_1 + n_2 + n_3}$$

pooled estimator

$$\left(\frac{\overset{f_1}{\cancel{n_1 \bar{X}_1}} - n_1 \hat{\theta}}{\sqrt{n_1 \hat{\theta}(1-\hat{\theta})}} \right)^2 + \left(\frac{f_2 - n_2 \hat{\theta}}{\sqrt{n_2 \hat{\theta}(1-\hat{\theta})}} \right)^2 + \left(\frac{f_3 - n_3 \hat{\theta}}{\sqrt{n_3 \hat{\theta}(1-\hat{\theta})}} \right)^2 \sim \chi^2_2$$

(7)

Regression analysis: $\mu_{Y|X} = \mathbb{E}(Y|X)$.

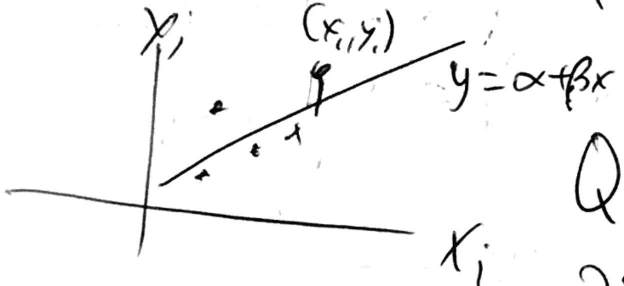
regression curve

If linear,

$$\mu_{Y|X} = \alpha + \beta X. \Rightarrow \mu_{Y|X} - \mathbb{E}(Y) = \frac{Cov(X,Y)}{Var(X)}(X - \mathbb{E}(X)).$$

Pearson correlation coefficient $\rho = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

Estimators for α & β given (x_i, y_i).



$$Q = \sum_{i=1}^{n} \left(y_i - (\alpha + \beta x_i)\right)^2 \leftarrow \text{minimize wrt } \alpha, \beta.$$

$$\frac{\partial Q}{\partial \alpha} = 0 = \frac{\partial Q}{\partial \beta}.$$

$$\hat{\alpha} = \overline{Y} - \overline{X} \cdot \frac{S_{xy}}{S_{xx}} \qquad S_{xy} = \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}).$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

If $f_Y(y|\underline{X=x})$ = normal. & regression linear

$$f_Y(y|X=x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-(\alpha+\beta x))^2}{2\sigma^2}}$$

Max likelihood $\Rightarrow$ estimator: $\hat{\alpha}, \hat{\beta}, \hat{\sigma}$

$\hat{\beta}$ as RV is normal $\mathbb{E}(\hat{\beta}) = \beta$.

$Var(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$. $\Rightarrow$ Confidence intervals & p-values for $\hat{\beta}$.

Is $\hat{\beta} = 0$?