Functions of a random variable (i.e. finding $f_Y(y)$ from $f_X(x)$ where $Y = g(X)$)

- generally, you find $F_X(x)$, then $F_Y(y) = P(Y \le y) = P(X \le g^{-1}(y))$

$= F_X(g^{-1}(y))$ if g is strictly monotonic increasing

Magic formula: If X is cts ranvar with density $f_X(x)$, s.t. $f_X(x) = 0$ if $x \notin I$ ← interval,

and g is differentiable & strictly monotonic on I,

then $Y = g(X)$ has density $f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|$

Joint distribution: $F(x,y) = P(X \le x, Y \le y)$

$$F(x_0, y_0) = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f(x,y)\, dy\, dx$$

$$f(x_0, y_0) = \frac{\partial^2}{\partial x \partial y} F(x_0, y_0)$$

$$f_X(x_0) = \int_{-\infty}^{+\infty} f(x_0, y)\, dy \quad \leftarrow \text{marginal distribution}$$

Independence: joint c.d.f. factors into product of marginal c.d.f.s:

$$F(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) \quad \text{for all } x_1, \dots, x_n$$

or equiv. the joint density function factors:

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) \quad \text{for all } x_1, \dots, x_n.$$

Quotient of cts ranvars: $f_{X,Y}(x,y)$ is given, and $Z = Y/X$

Then $f_Z(z) = \int_{-\infty}^{+\infty} |x|\, f(x, xz)\, dx$

- if X & Y are indep, then $f_Z(z) = \int_{-\infty}^{+\infty} |x|\, f_X(x)\, f_Y(xz)\, dx$

Extrema

$V \leftarrow \min_i X_i$, $U \leftarrow \max_i X_i$ (x_1, x_3, x_2, x_4)

where X_i are indep. and each have c.d.f. F and density f.

- $F_U(u) = P(U \le u) = P(X_1 \le u) \cdots P(X_n \le u) = F(u)^n$

$\therefore f_U(u) = \frac{d}{du} F(u)^n = n f(u) F(u)^{n-1}$

- $F_V(v) = 1 - (1 - F_V(v))^n$

$\therefore f_V(v) = n f(v) (1 - F(v))^{n-1}$

- E.g. if $X_i \sim Exp(\lambda)$, then $f_V(v) = n \cdot \lambda e^{-\lambda v} \cdot (e^{-\lambda v})^{n-1} = n\lambda e^{-n\lambda v}$
- E.g. if $X_i \sim Uniform(0, \theta)$, then $f_U(u) = n \cdot f(u) \cdot F(u)^{n-1} = \frac{n u^{n-1}}{\theta^n}$ (where $0 \le u \le \theta$)

Expectation $E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx$

- of a function $Y = g(X)$: $E(Y) = \int_{-\infty}^{+\infty} g(x) f(x)\, dx$
- of a function over jointly distrib. ranvar $Y = g(X_1, \dots, X_n)$: $E(Y) = \int \cdots \int g(x_1, \dots, x_n) f(x_1, \dots, x_n)\, dx_1 \cdots dx_n$
- if X, Y indep. ranvar : $E[XY] = E[X] \cdot E[Y]$
- if X, Y indep. ranvar & g, h are any functions, then $E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$
- linearity of expectation: $E(a + \sum_{i=1}^{n} b_i X_i) = a + \sum_{i=1}^{n} b_i E(X_i)$

Variance $Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$

- $Var(a + bX) = b^2 Var(X)$
- Chebyshev's ineq : $P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$
  (if σ^2 is small then X should not deviate too much from μ)
- If X_i are indep, then $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$

Moment generating function : $M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$ (cts)

$= \sum_i e^{tx_i} p(x_i)$ (discrete)

- m.g.f. uniquely determines p.d.f. if m.g.f. exists for t in an open interval containing zero
- $M^{(r)}(0) = E[X^r]$

  rth differentiation
- $M_{a+bX}(t) = e^{at} M_X(bt)$ (moment of a linear change of variable)
- if X, Y indep and $Z = X + Y$, then $M_Z(t) = M_X(t) M_Y(t)$ (on the common interval where M_X and M_Y exist)

Limit theorems

- **Weak Law of Large Numbers (WLLN)**:

  If $X_1, \dots, X_i, \dots$ are a sequence of indep. ranvar with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ (not necessarily normal),

  Then $\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$ satisfies:

  $$\forall \varepsilon > 0, \quad P(|\bar{X}_n - \mu| > \varepsilon) \to 0 \text{ as } n \to \infty$$ (Pf: using Chebyshev)

  (in other words, taking more indep. readings gets us arbitrarily close to the mean)
- "Converge in probability to α" : $P(|Z_n - \alpha| > \varepsilon) \to 0$ as $n \to \infty$
- If we don't know σ^2, we can estimate it with $\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2$

- "Converges in distribution":

If $X, X_1, X_2, \dots$ is a sequence of ranvars with c.d.f.s $F, F_1, F_2, \dots$ resp, then

X_n converges in distribution to X if $\underbrace{F_n(x) \to F(x)}_{\text{pt wise convergence of the c.d.f.}}$ as $n \to \infty$

- Central Limit Theorem (CLT):

If $X_1, X_2, \dots$ are indep ranvars with mean μ and variance σ^2 with c.d.f. F (and m.g.f. is defined in a neighbourhood of 0),

then: $S_n := \sum_{i=1}^{n} X_i$

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \to \underbrace{\Phi(x)}_{\text{std. normal c.d.f.}} \text{ as } n \to \infty$$

(i.e. $\frac{S_n}{n}$ fluctuates around μ in a normal dist. with variance $\frac{\sigma}{\sqrt{n}}$)

- skewed dist. or with large tails will need larger n for a good approx.
- sometimes, we are given $E(\hat{\alpha})$ and $Var(\hat{\alpha})$ that are already obtained as an estimate from a large sample. Then,

$$P\left(\frac{\hat{\alpha} - E(\hat{\alpha})}{\sqrt{Var(\hat{\alpha})}} \leq x\right) \to \Phi(x) \text{ as } n \to \infty$$

(or in other words, $\hat{\alpha} \to N(E(\hat{\alpha}), Var(\hat{\alpha}))$)

t Distribution : Given $Z \sim N(0,1)$ (i.e. std normal) and $U \sim \chi^2_n$ $\}$ indep

then $T = \dfrac{Z}{\sqrt{U/n}} \sim t_n$ ← t distribution with n d.o.f.

(t_1 is "Cauchy distribution")

- $f(t) = \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1+\dfrac{t^2}{n}\right)^{-\frac{n+1}{2}}$
- note: it is symmetric about 0, since $f(t) = f(-t)$
- As $n \to \infty$, $t_n \to N(0,1)$ (tails become lighter)

F Distribution : Given $U \sim \chi^2_m$ and $V \sim \chi^2_n$ $\}$ indep

then $W = \dfrac{U/m}{V/n} \sim F_{m,n}$ ← F distribution with m and n degrees of freedom.

- $f(w) = \dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\dfrac{m}{n}\right)^{\frac{m}{2}} w^{\frac{m}{2}-1} \left(1+\dfrac{m}{n}w\right)^{-\frac{m+n}{2}}$
- $E(W) = \dfrac{n}{n-2}$ for $n > 2$
- $(t_n)^2 \sim F_{1,n}$
- If $\underbrace{X, Y}_{\text{indep}} \sim \text{Exp}(\lambda=1)$, then $\dfrac{X}{Y} \sim F_{2,2}$

On normal distributions & sample variance

- If $X \sim N(\mu_X, \sigma^2_X)$ and $Y \sim N(\mu_Y, \sigma^2_Y)$ (indep) then $X+Y \sim N(\mu_X + \mu_Y, \sigma^2_X + \sigma^2_Y)$
- Let $X_1, \ldots, X_n$ be indep $N(\mu, \sigma^2)$ ranvars:
  - $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (sample mean) and $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
  - $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ (sample variance)

    note the $(n-1)$.
  - $\bar{X}$ and $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ are indep.
  - $\bar{X}$ and S^2 are indep.
  - $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ (or $\dfrac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$, where $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - X)^2$)
  - $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

Estimation

- given data points $X_1, \dots, X_n$ (indep, same dist)
- want to estimate θ (parameters for dist) ; the indiv. distrib is $f(x|\theta)$, joint distrib is $f(x_1|\theta) \dots f(x_n|\theta)$

Method of Moments

- k^{th} moment is $\mu_k := E(X^k)$
- want observed $\hat{\mu}_k$ to be equal to expected μ_k.
- $\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$
- Steps: (1) Express low-order moments (starting from $k=1$) in terms of parameters
  if moment does not depend on params, then ignore.
  e.g. μ, σ^2
  (e.g. $\mu_1 = E(X) = \mu$; $\mu_2 = E(X^2) = \mu^2 + \sigma^2$)
  (2) Rearrange expressions to make parameters the subjects
  (e.g. $\mu = \mu_1$; $\sigma^2 = \mu_2 - \mu_1^2$)
  (3) Calculate and substitute sample moments
  (e.g. calculate $\hat{\mu}_1, \hat{\mu}_2$ from the given data, sub in μ_1, μ_2 to find μ and σ^2)
  e.g. $\hat{\mu} = \bar{X}$; $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$

Bias:

- How much $E(\hat{\theta})$ differs from θ

  expectation of guessed param — true param.

(e.g. $E(\hat{\alpha}) \underset{\text{from MoM estimate}}{=} 3E[\bar{X}] \underset{\text{L.o.E.}}{=} 3 \frac{1}{n} \sum_i E[X_i] \underset{\text{from the mean (based on distribution param)}}{=} 3 \frac{1}{n} \cdot n \cdot \frac{\alpha}{3} = \alpha$)

Consistency of an estimate $\hat{\theta}_n$ of θ

computed using n samples.

- $\hat{\theta}_n$ is consistent in probability if $\hat{\theta}_n$ converges in probability to θ as $n \to \infty$
  i.e. $\forall \varepsilon > 0$, $P(|\hat{\theta}_n - \theta| > \varepsilon) \to 0$ as $n \to \infty$.
- MoM estimates are consistent (due to WLLN)
  - so (assuming functions are continuous), estimates will converge.

Method of Maximum Likelihood

- Random vars $X_1, \dots, X_n$; parameter to estimate: θ.
- joint density or freq: $f(x_1, \dots, x_n | \theta) = \underbrace{f(x_1|\theta) \dots f(x_n|\theta)}_{\text{if indep. \& i.d.}}$
- $\text{lik}(\theta) = f(x_1, \dots, x_n | \theta)$
  $= \prod_{i=1}^{n} f(x_i | \theta)$ (if i.i.d.)

  want to find θ that maximises this.
- equiv. to finding θ that maximises $l(\theta) := \log(\text{lik}(\theta)) = \overbrace{\sum_{i=1}^{n} \log(f(x_i|\theta))}^{\text{if i.i.d.}}$
  - differentiate $l(\theta)$ ... use partial derivative if more than one parameter

(page 35 for e.g.'s)

MLE of Multinomial cell prob. → p. 36.

Large sample theory for MLE (p.38)

- Fischer information: $I(\theta) = \underset{\substack{\uparrow \\ \text{over } X \\ \text{given } \theta}}{E}\left(\left[\frac{\partial}{\partial\theta}\log f(x|\theta)\right]^2\right)$ (indiv. sample)   natural log
  - if sufficiently smooth: $I(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right)$
    may not hold if support of f depends on θ.
  - if sufficiently smooth, and i.i.d.
    (a) MLE $\hat{\theta}$ is consistent
    (b) $\boxed{\sqrt{nI(\theta_0)}\,(\underset{\text{estimate}}{\hat{\theta}} - \underset{\text{true value}}{\theta_0}) \to N(0,1)}$ (i.e. $\hat{\theta}$ is approx $N\left(\theta_0, \underbrace{\frac{1}{nI(\theta_0)}}_{\text{asymptotic variance}}\right)$)

Fischer info of i.i.d. sample of size n is $nI(\theta)$

- $\boxed{\text{If } I(\theta_0) \text{ is unknown, use } I(\hat{\theta}) \text{ instead.}}$
- If not i.i.d., then Fisher information of sample is
$$E[\ell'(\theta)^2] \quad \text{or} \quad -E[\ell''(\theta)]$$
and the asymptotic variance is $\frac{1}{E[\ell'(\theta)^2]}$ or $-\frac{1}{E[\ell''(\theta)]}$ (over X given θ)

Confidence interval:

$100(1-\alpha)\%$ confidence interval for θ → interval that contains θ with probability $1-\alpha$.

- For normal dist:
  - $100(1-\alpha)\%$ confidence interval for μ is $\bar{X} \pm \frac{S}{\sqrt{n}} t_{n-1}\left(\frac{\alpha}{2}\right)$ (S: sample s.d.)   (since $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$)
  - $100(1-\alpha)\%$ confidence interval for σ^2 is $\left(\frac{n\hat{\sigma}^2}{\chi^2_{n-1}\left(\frac{\alpha}{2}\right)}, \frac{n\hat{\sigma}^2}{\chi^2_{n-1}\left(1-\frac{\alpha}{2}\right)}\right)$ ← right tail cumulative   (since $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$)
- Large sample theory approx:
$$P\left(-z\left(\frac{\alpha}{2}\right) \le \sqrt{nI(\hat{\theta})}\,(\hat{\theta}-\theta_0) \le z\left(\frac{\alpha}{2}\right)\right) \approx 1-\alpha$$
← right tail cumulative

so $100(1-\alpha)\%$ confidence interval for θ is $\hat{\theta} \pm \frac{z\left(\frac{\alpha}{2}\right)}{\sqrt{nI(\hat{\theta})}}$.

Efficiency

- Mean squared error (MSE): $E[(\hat{\theta}-\theta_0)^2] = \underbrace{Var(\hat{\theta})}_{\text{variance}} + \underbrace{(E(\hat{\theta})-\theta_0)^2}_{\text{squared bias}}$
- Efficiency of $\hat{\theta}$ relative to $\tilde{\theta}$: $eff(\hat{\theta},\tilde{\theta}) = \dfrac{Var(\tilde{\theta})}{Var(\hat{\theta})}$ ← how concentrated the estimate is

  two estimates of θ

- Cramér-Rao lower bound: for any unbiased estimate of θ (from n i.i.d. ranvars)

$$Var(T) \geq \frac{1}{nI(\theta)}$$

  - gives the best possible variance
  - An efficient unbiased estimate attains this lower bound
    - if it is biased, then we don't say that it is efficient.

Sufficiency

Sufficiency: finding a statistic that contains all the info in a sample about θ

- A statistic $T=T(X_1,\dots,X_n)$ is sufficient for θ if $P(X_1=x_1,\dots,X_n=x_n \mid T=t)$ does not depend on θ (for any t)
- ** Equiv: $T(\tilde{X})$ is sufficient for θ iff $f(\tilde{x}\mid\theta) = g(T(\tilde{x}),\theta)\cdot h(\tilde{x})$

  (i.e. f factorizes into g and h)

  e.g. for $X_i \sim N(\mu,\sigma^2)$,

$$f(\tilde{x}\mid\mu,\sigma) = \frac{1}{\sigma^n(2\pi)^{\frac{n}{2}}}\exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2\right)\right)$$

  so $\sum_{i=1}^n x_i$ and $\sum_{i=1}^n x_i^2$ are (together) sufficient statistics

- Exponential family of dists: those with one-parameter → $f(x\mid\theta) = \begin{cases} \exp(c(\theta)T(x)+d(\theta)+S(x)), & x\in A \\ 0, & x\notin A \end{cases}$

  (θ can be a vector)

  where A does not depend on θ.

  (i.i.d.) $f(\tilde{x}\mid\theta) = \prod_{i=1}^n f(x\mid\theta)$ → These factorize into $\exp\left(c(\theta)\sum_{i=1}^n T(x_i) + n\,d(\theta)\right)\exp\left(\sum_{i=1}^n S(x_i)\right)$

  so $\sum_{i=1}^n T(x_i)$ is a sufficient statistic.

  k-parameters: $f(x\mid\theta) = \begin{cases} \exp\left[\sum_{j=1}^k c_j(\theta)T_j(x) + d(\theta) + S(x)\right], & x\in A \\ 0, & x\notin A \end{cases}$

  The parameters $\sum_{i=1}^n T_j(x_i)$ $(1\leq j\leq k)$ are sufficient statistics

Rao-Blackwell thm : (sufficient statistics make the best estimators)

- If $\hat{\theta}$ is an estimator of θ with $E(\hat{\theta}^2) < \infty$ for all θ,

  and T is sufficient for θ,

  and $\tilde{\theta} = E(\hat{\theta} \mid T)$,

  then $\forall\, \theta$, $E[(\tilde{\theta}-\theta)^2] \leq E[(\hat{\theta}-\theta)^2]$

  (where equality holds iff $\hat{\theta} = \tilde{\theta}$)

Hypothesis Testing

- H_0 : null hypothesis
- H_1 : alternative hypothesis
- Type I error : rejecting H_0 when it is true
- $P(\text{Type I error}) = \alpha$: significance level of the test
- Type II error : accepting H_0 when it is false
- $P(\text{Type II error}) = \beta$
- $P(H_0 \text{ rejected when it is false}) = 1-\beta$: power of the test
- test statistic : some value X s.t. we decide to accept/reject H_0 based on X.
- acceptance region : set of values for test statistic s.t. we accept H_0
- rejection region : set of values for test statistic s.t. we reject H_0
- null distribution : distribution of test statistic under H_0
- simple hypothesis : H_0 and H_1 are completely specified distributions (i.e. no unfixed variables)
  (e.g. $H_0 : X \sim \text{binomial}(10, 0.5)$)

Neyman-Pearson lemma : Given that H_0 and H_1 are simple hypotheses, for any significance level α, the likelihood ratio test has the most power.

Likelihood ratio test : test statistic $= \dfrac{P(x \mid H_0)}{P(x \mid H_1)}$ → Want to reject H_0 when likelihood ratio $< c$ (for some c to get the desired significance level α)

likelihood ratio

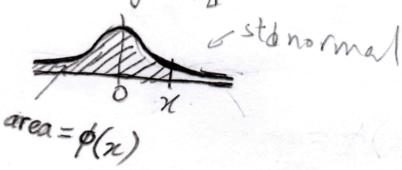
x is a vector of all the samples

- usually, we can map this back to something like $\bar{X} > x_0$ or $\bar{X} < x_0$.
- see p. 49 for example when α is given

p-value : smallest significance level under which H_0 will be rejected
(i.e. $P(T \geq t_{obs} \mid H_0)$)

depends on direction

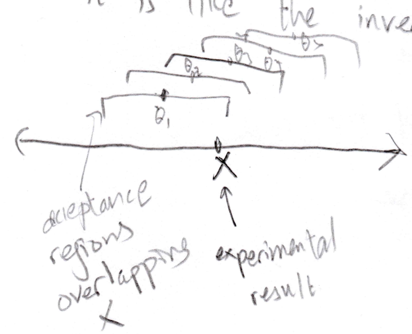
the observed value



Uniformly most powerful test : The rejection region depends on the null distribution and H_0 but not on H_1's variables (e.g. μ_1) apart from the direction

Confidence intervals: p. 52, Ex 34.

:= the set of values for μ_0 for which $H_0: \mu = \mu_0$ is accepted.
(so it is like the inverse of a hypothesis test)



- confidence interval is the set of all θ where X is accepted.

- $C(X) = \{\theta \mid X \in A(\theta)\}$

$C(X)$: $100(1-\alpha)\%$ confidence region for θ.

$A(\theta)$: acceptance region of θ at level α

- $A(\theta_0) = \{X \mid \theta_0 \in C(X)\}$

— p. 53, Ex 35.

Generalized likelihood ratio tests for non-simple (i.e. composite) hypotheses:

- not generally optimal, but one of the best ways since no optimal tests exist.
- Let Ω be the set of all possible values of θ, and $\omega_0 \subseteq \Omega$ be the subset where H_0 is valid (i.e. $H_0: \theta \in \omega_0$)

Then $\Lambda = \dfrac{\max_{\theta \in \omega_0} \text{lik}(\theta)}{\max_{\theta \in \Omega} \text{lik}(\theta)}$

rejection region: $\Lambda \leq \lambda_0$ (i.e. it is not likely that $\theta \in \omega_0$)
for some λ_0.

- $-2\log\Lambda \longrightarrow \chi^2$ with $df = \dim\Omega - \dim\omega_0$

always natural

df: degrees of freedom

where $\dim S$ is the number of free params under S.

- e.g. for $H_0: \mu = \mu_0$ and $H_1: \mu \neq \mu_0$ where $X_1, \ldots, X_n$ i.i.d. from normal dist with variance σ^2,
then $\dim\Omega = 1$ (μ is free ~~variable~~ but not σ^2), $\dim\omega_0 = 0$ (no free params),
so $-2\log\Lambda \to \chi_1^2$.
(actually, it is equal for normal dist)

Likelihood ratio tests for ~~multi~~ multinomial distribution (p.55, Ex. 36.)

- the vector $\tilde{p} = \tilde{p}(\theta)$ where $\theta \in \omega_0$ for some θ
- for Ω, the $\tilde{p}$ are free such that they are valid (i.e. $p_i \geq 0$ and $\sum_i p_i = 1$

· likelihood ratio $\Lambda = \dfrac{\max_{\theta \in \omega_0} \text{lik}(\theta)}{\max_{\theta \in \Omega} \text{lik}(\theta)}$ $\rightarrow \max_{\theta \in \omega_0} \left(\dfrac{n!}{x_1! \cdots x_m!}\right) p_1(\theta)^{x_1} \cdots p_m(\theta)^{x_m}$

where $x_1, \ldots, x_m$ are observed cell counts

$$= \frac{\left(\frac{n!}{x_1! \cdots x_m!}\right) p_1(\hat{\theta})^{x_1} \cdots p_m(\hat{\theta})^{x_m}}{\left(\frac{n!}{x_1! \cdots x_m!}\right) \hat{p}_1^{x_1} \cdots \hat{p}_m^{x_m}}$$

where $\hat{\theta}$ is the mle

and $\hat{p}_i = \frac{x_i}{n}$ is the unrestricted mle

$$= \prod_{i=1}^{m} \left(\frac{p_i(\hat{\theta})}{\hat{p}_i}\right)^{x_i}$$

$$\therefore -2\log\Lambda = 2\sum_{i=1}^{m} O_i \log\left(\frac{O_i}{E_i}\right)$$ where $O_i = n\hat{p}_i = x_i$ (observed count)

$E_i = np_i(\hat{\theta})$ (expected count)

$\dim \Omega = m-1$ (since there is a constraint that $\sum_i p_i = 1$)

$k := \dim \omega_0$ is the dimensionality of θ (possibly a vector)

So $-2\log\Lambda \rightarrow \chi^2_{m-k-1}$

Pearson's X^2 statistic: $X^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}$

(Ex. 36)

$$X^2 \longrightarrow -2\log\Lambda \longrightarrow \chi^2_{m-k-1}$$

not equal, but asymptotically equivalent under H_0

Comparing two samples

If $X_1, \dots, X_n \overset{iid}{\sim} N(\mu_X, \sigma^2)$

and $Y_1, \dots, Y_m \overset{i.i.d.}{\sim} N(\mu_Y, \sigma^2)$

we want to see if $\mu_X = \mu_Y$ or not.

- If σ^2 is known, then $$Z = \frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim N(0,1)$$

  - So a $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is $(\bar{X}-\bar{Y}) \pm z(\frac{\alpha}{2})\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}$

- If σ^2 is unknown, then estimate it from pooled sample variance:

$$s_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{m+n-2}$$

where $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

and $S_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2$

$$s_{\bar{X}-\bar{Y}} = s_p\sqrt{\frac{1}{n}+\frac{1}{m}} \quad \text{(est. std. error of } \bar{X}-\bar{Y})$$

note: no square

$$t = \frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{s_{\bar{X}-\bar{Y}}} \sim t_{m+n-2}$$

t distribution with m+n-2 deg. of freedom

- so a $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is $(\bar{X}-\bar{Y}) \pm t_{m+n-2}(\frac{\alpha}{2})\, s_{\bar{X}-\bar{Y}}$

- For hypothesis testing:
  - test statistic $t = \frac{\bar{X}-\bar{Y}}{s_{\bar{X}-\bar{Y}}}$
  - $H_0: \mu_X = \mu_Y$
  - $H_1: \mu_X \neq \mu_Y \iff |t| > t_{m+n-2}(\frac{\alpha}{2})$
  - $H_2: \mu_X > \mu_Y \iff t > t_{m+n-2}(\alpha)$
  - $H_3: \mu_X < \mu_Y \iff t < -t_{m+n-2}(\alpha)$

- If variances of X_i's and Y_i's are not equal, then: $t = \frac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}}$ (?)

  - test statistic $t = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}}$ $\sim t_{df}$

    where $df = \frac{\left(\frac{S_X^2}{n}+\frac{S_Y^2}{m}\right)^2}{\left(\frac{S_X^2}{n}\right)^2/(n-1) + \left(\frac{S_Y^2}{m}\right)^2/(m-1)}$

  test t

- nonnormality: For large sample size, it is approximately valid