Homework #1

Q1: Sps that $X_1, X_2$ are iid r.v.'s and that pdfs are $f(x) = \begin{cases} e^{-x}, & x>0 \\ 0, & \text{o.w} \end{cases}$

a) What distribution/family do $X_1, X_2$ follow?

Ans: Exponential since $f(x) = \lambda e^{-\lambda x}$ for $x>0$, follows when $\lambda = 1 \Rightarrow f(x) = e^{-x} : x>0$

Know that by independence $f(x_1, x_2) = f(x_1) f(x_2) = e^{-x_1} e^{-x_2} = e^{-(x_1+x_2)}$

b) Find expectation & variance of $y = X_1 - X_2$

Ans: Want to find $E(y) = E(x_1 - x_2)$. Know that $E(X) = \int x f(x) dx \Rightarrow$ $E(X_1) = \int_0^\infty x_1 e^{-x_1} dx_1$, $E(X_2) = \int_0^\infty x_2 e^{-x_2} dx_2$

By kernel method we know that $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{t} dt$, and $\Gamma(\alpha) = (\alpha-1)!$

$\Rightarrow E(X_1) = \int_0^\infty x_1 e^{-x_1} dx = \int_0^\infty x_1^{2-1} e^{-x_1} dx_1 = \frac{\Gamma(2)}{(1)^2} = \frac{(2-1)!}{1} = 1$, same for $x_2$.

$\therefore E(X_1 - X_2) = E(X_1) - E(X_2) = 1 - 1 = 0$, $E(X_1^2) = \int_0^\infty x_1^2 e^{-x_1} dx_1 = \int_0^\infty x_1^{3-1} e^{-x_1} dx_1 = \frac{\Gamma(3)}{1^2} = \frac{(3-1)!}{1} = 2$

$\Rightarrow var(x_1) = E(X_1^2) - E(x_1)^2 = 2 - 1 = 1$, same for $x_2$.

$\Rightarrow var(y) = var(x_1 - x_2) = var(X_1) + var(X_2) - 2cov(X_1, X_2)$

and $cov(x_1, x_2) = E(x_1 x_2) - E(x_1)E(x_2)$ but $E(X_1 X_2) = E(X_1)E(X_2)$ by independence

$\Rightarrow cov(x_1, x_2) = 0$ and $var(y) = var(X_1) + var(X_2) = 1 + 1 = 2$*

$E(y) = E(x_1 - x_2) = E(x_1) - E(x_2) = 1 - 1 = 0$*

Q2: Sps $x_1, x_2, x_3$ are r.v's and independently distributed from Bernoulli with parameter $p = 0.5$

a) Find $E(2x_1 + x_3 - 4)$ Ans: $= 2E(X_1) + E(x_3) - 4 = 2(\frac{1}{2}) + \frac{1}{2} - 4 = \frac{-5}{2}$

b) $E[(x_1 - 2x_2 + x_3)^2] = E[x_1^2 + 4x_2^2 + x_3^2 - 4x_1x_2 + 2x_1x_3 - 4x_2x_3]$

Ans: $[(a+b+c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc]$ $= E(x_1^2) + 4E(x_1x_2) + 2E(x_1x_3) - 4E(x_2x_3)$

$= \frac{1}{2} + 4(\frac{1}{2}) + \frac{1}{2} - 4E(x_1)E(x_2) + 2E(x_1)E(x_3) - 4E(x_2)E(x)$

$\Rightarrow 3 + 4(\frac{1}{2}) + 2(\frac{1}{2}) - 4(\frac{1}{2}) = \frac{3}{2}$

Q3: Sps. X is distributed from a Norm. $N(\mu, \sigma^2)$ with $\mu = 2, \sigma^2 = 1$ Define $y = 2x - 1$

Ans: we know $E(y) = aE(x) + b$, $var(y) = a^2 var(x) \Rightarrow E(y) = 2E(x) - 1 = 2(2) - 1 = 3$

and $var(y) = 2^2 var(X) = 4(1)$

b) find prob: $P(1 < y < 3) = P(1 < 2x - 1 < 3) = P(1 < x < 2) \Rightarrow normcdf(1, 2, \mu = 2, \sigma^2 = 1)$ $\approx 0.34134$

Q4: Show that if $X_{11}, \ldots, X_{1n_1}$ and $X_{21}, \ldots X_{2n_2}$ are independent random variables with first $n_1$ constituting a random sample from infinite population with mean $\mu_1$ and variance $\sigma_1^2$, same for second random sample

a) $E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$ Ans: $E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2)$,

and know $\bar{X}_1 = \sum_{i=1}^{n_1} X_{1i}$, $\bar{X}_2 = \sum_{i=1}^{n_2} X_{2i}$, for arbitrary r.v. Y with pdf, f(y)

we know $E(y) = \sum_y y f(y) \Rightarrow E(\bar{X}_1) = E(\sum_{i=1}^{n_1} X_{1i}) = \mu_1$

$\therefore E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2)$ $E(\bar{X}_2) = E(\sum_{i=1}^{n_2} X_{2i}) = \mu_2$

$= \mu_1 - \mu_2$

b) $var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ Ans: $var(\bar{X}_1) = \sigma_1^2$, $var(\bar{X}_2) = \sigma_2^2$ independent

Thus $cov(\bar{X}_1, \bar{X}_2) = 0$, $var(\bar{X}_1 - \bar{X}_2) = var(\bar{X}_1) + var(x_2) - \cancel{2cov(X_1, X_2)}^{0}$

The variance of the sample mean is $= \sigma_1^2 + \sigma_2^2$

$var(\bar{X}_i) = \frac{\sigma_i}{n_i}$ $\forall i \in \{1, 2\}$ $\Rightarrow var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

Q5: If first $n_1$ r.v's from Q4 have Bernoulli dist with param $\theta_1$, and the other $n_2$ have $\theta_2$, show that for the sample proportions $\hat{\theta}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i}$, $\hat{\theta}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} x_{2i}$

a) $E(\hat{\theta}_1 - \hat{\theta}_2) = \theta_1 - \theta_2$ Ans: We know that $X_{1i}, X_{ii} \sim Bernoulli(\theta_i)$ respectively

$\Rightarrow E(x_{1i}) = \theta_1$, $var(x_{1i}) = \theta(1-\theta)$ $\because$ Bernoulli, same for $x_{2i}$

$\Rightarrow \tilde{\theta}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ $\forall i = \{1, 2\}$ $\Rightarrow E(\hat{\theta}_1 - \hat{\theta}_2) = E(\sum_{i=1}^{n_1} x_{1i} - \sum_{i=1}^{n_2} x_{2i}) = \frac{1}{n_1}E(\sum_{i=1}^{n_1} x_{1i}) - \frac{1}{n_2}E(\sum_{i=1}^{n_2} x_{2i})$

$\Rightarrow \frac{1}{n_1} n_1 \theta_1 - \frac{1}{n_2} n_2 \theta_2 = \theta_1 - \theta_2$

b) $var(\hat{\theta}_1 - \hat{\theta}_2) = \frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}$ Ans: $var(\hat{\theta}_1 - \hat{\theta}_2) = var(\frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i}) - var(\frac{1}{n_2}\sum_{i=1}^{n_2} x_{2i})$

$\Rightarrow \frac{1}{n_1^2} var(\sum_{i=1}^{n_1} x_{1i}) - \frac{1}{n_2^2} var(\sum_{i=1}^{n_2} x_{2i}) \Rightarrow \frac{1}{n_1^2} n_1 var(x_{1i}) - \frac{1}{n_2^2} n_2 var(x_{2i})$

$\Rightarrow \frac{1}{n_1} var(x_{1i}) - \frac{1}{n_2} var(x_{2i}) \Rightarrow \frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}$

Homework #2:

Q1: ... Sps we model the concentration of lactic acid in chunks as independent normal r.v's w/ mean $\mu$ and variance $\sigma^2$ (both $\mu, \sigma^2$ known) Interested in how much concentrations differ from $\mu$. let $X_1, \ldots, X_k$ be concentrations of k chunks and $Z = \frac{X_i - \mu}{\sigma}$

then $y = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2 = \frac{\sigma^2}{k}\sum_{i=1}^{k} z_i^2$ is one measure of how much the k conc. differ from $\mu$.

a) name dist of $W = \frac{k y^2}{\sigma^2}$ Ans: $W = \frac{k}{\sigma^2}\left(\frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2\right) = \frac{k}{\sigma^2}\left(\frac{\sigma^2}{k}\sum_{i=1}^{k} z_i^2\right) = \sum_{i=1}^{k} z_i^2$

W is the sum of k independent chunks. It follows that $W \sim \chi^2_k$

b) Sps $k = 10, \sigma^2 = 0.09$ find c: $P(y \leq c) = 0.99$

Ans: $P(y \leq c) = 0.99 \Rightarrow P(W \leq \frac{k}{\sigma^2} c) = 0.99 \Rightarrow$ find 99th percentile of $\chi^2_{10}$

$qchisq(.99, 10) = 23.20925 \Rightarrow \frac{k}{\sigma^2} c = 23.20925 \Rightarrow c = 0.208881$

Q2: In Q1 now sps $k = 20$ chunks, $X_1, \ldots, X_{20}$ let $\bar{X}_{20} = \frac{1}{k}\sum_{i=1}^{20} x_i$. Find c: $P(\bar{X}_{20} \leq \mu + c\sigma) = 0.95$

Ans: $P(\bar{X}_{20} \leq \mu + c\sigma) = 0.95 \Rightarrow P(Z = \frac{\bar{X}_{20} - \mu}{\sigma/\sqrt{20}} \leq \frac{(\mu + c\sigma) - \mu}{\sigma/\sqrt{20}} = z) = 0.95$

$\Rightarrow P(Z \leq c\sqrt{20} = z) = 0.95$ find critical value

$1.64485 = invNorm(area = 0.95, 0, 1, LEFT) = 1.64485 = c\sqrt{20} \Rightarrow c \approx 0.3678$

Q3: verify: $\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 - n(\bar{x} - \mu)^2$

LHS: $\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \mu + \bar{x} - \bar{x})^2 = \sum_{i=1}^{n}[(x_i - \bar{x}) + (\bar{x} - \mu)]^2$

$\Rightarrow \sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(\bar{x} - \mu)^2 + 2\sum_{i=1}^{n}(x_i - \bar{x})(\bar{x} - \mu)$ factor out

$\bar{x} = \frac{1}{n}\sum x_i$

$\Rightarrow \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 + 2(\bar{x} - \mu)\sum_{i=1}^{n}(x_i - \bar{x})$

$\Rightarrow \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$ $\cancel{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i}^{0}$

Q4: A random sample of $n = 100$ is taken from infinite pop w/ mean $\mu = 75$ and variance $\sigma^2 = 256$

a) Based on LLN w/ what prob. can we assert the value we obtain for $\bar{X}$ will fall between 67 and 83? Ans: $\mu = 75$ lower bound $= 75 - 67 = 8$, upper bound $= 83 - 75 = 8$ $\Rightarrow \varepsilon = 8$

we know that $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{256}}{\sqrt{100}} = \frac{16}{10} = 1.6$

By LLN: $P(|X - \mu| \leq \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$ (LLN) $\Rightarrow P(|\bar{X} - 75| \leq 8) \geq 1 - \frac{256}{100(8)^2}$

$\Rightarrow P(-8 \leq \bar{X} - 75 \leq 8) \geq 1 - \frac{256}{6400} \Rightarrow P(67 \leq \bar{X} \leq 83) \geq 1 - 0.04$

3) By Chebyshev's $k = \frac{83 - \mu}{\sigma_{\bar{X}}} \Rightarrow \frac{83 - 75}{1.6} = 5$ $P(67 \leq \bar{X} \leq 83) \geq 0.96$

By Chebyshev's $\Rightarrow P(|\bar{X} - \mu| > k\sigma) \leq \frac{1}{k^2} \Rightarrow P(|\bar{X} - \mu| \leq k\sigma_{\bar{X}}) \geq 1 - \frac{1}{k^2} \Rightarrow P(|\bar{X} - 75| \leq 8) \geq 1 - 0$

b) By CLT what prob can we assert that the value we obtain for $\bar{X}$ will fall between 67 and 83? $\hookrightarrow P(67 \leq \bar{X} \leq 83) \geq 0.4$

Ans: $P(67 \leq \bar{X} \leq 83) = P(\frac{67 - \mu}{\sigma_{\bar{X}}} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq \frac{83 - \mu}{\sigma_{\bar{X}}}) = P(\frac{67 - 75}{1.6} \leq Z \leq \frac{83 - 75}{1.6})$

$\Rightarrow P(-5 \leq Z \leq 5) = \Phi(5) - \Phi(-5) = .9999994 \approx 1$

Q5: A random sample of $n = 225$ is taken from a uniform population with $\alpha = 24, \beta = 48$ .04

Based on CLT what's the prob. that the mean of the sample will be less than 35? 6

$X \sim Unif[\alpha, \beta]$, $E(X) = \frac{\alpha + \beta}{2}$, $var(X) = \frac{(\beta - \alpha)^2}{12} \Rightarrow$ apply CLT.

Ans: $P(\mu < 35)$: $\mu = E(X) = \frac{24 + 48}{2} = 36$, $\sigma = var(x) = \frac{(\beta - \alpha)^2}{12} = \frac{24^2}{12} = 48$

and $\sigma = \sqrt{\sigma^2} = \sqrt{48} = 6.9282$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{6.9282}{\sqrt{225}} = 0.46188$

Apply CLT: $P(Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq \frac{35 - 36}{\sigma_{\bar{X}}}) = P(Z < -\frac{1}{0.46188}) = 0.01519384$ $P(|X - \mu| < k\sigma) \leq$

Notes & Theorems: Chebyshev's: r.v. X will take on a value $k > 0$ std.dev. away from $\mu$

ex) Take on value between 0.95? $P(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2} \geq .95$

Thm 1: $X_1, \ldots, X_n$ form n Bernoulli trials with p, $X = X_1 + \ldots + X_n$ $\frac{1}{k^2}$

$\Rightarrow X \sim Binomial(n, p)$

Thm 2: $X_1, X_2$ independent $\sim$ Binomial with $n_1, n_2$ and p.

$\Rightarrow X_1 + X_2 \sim Binom(n = n_1 + n_2, p)$

Thm 3: $X_1, \ldots, X_n$ constitute a random sample w/ $\infty$ pop.

$\Rightarrow E(\bar{X}) = \mu$, $Var(\bar{X}) = \frac{\sigma^2}{n}$

IID: $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

1) $X_1, \ldots, X_n$ mutually independent $f(x_1, \ldots x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$

2) $X_1, \ldots, X_n$ are identically distributed from $N(\mu, \sigma^2)$

$f_{X_1}(x_1) = f_{X_2}(x_2) = \ldots = f_{X_n}(x_n) = N(\mu, \sigma^2)$

Function of R.V.'s $T: \mathbb{R}^n \to \mathbb{R}$, $T(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

$\Rightarrow \bar{X} = T(X_1, \ldots X_n)$ are r.v's.

Thm: $X_1, \ldots X_n$ be a sample of a r.v. X, let $T = T(X_1, \ldots X_n)$ be a function of the sample. Then T is called a statistic. (dist = sampling dist)

Thm (LLN) $\varepsilon > 0$, prob. of $\bar{X}$ will take on a value within $[\mu - \varepsilon, \mu + \varepsilon]$ is at least $1 - \frac{\sigma^2}{(n\varepsilon)^2}$, $P(|\bar{X} - \mu| \leq \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$

when $n \to \infty$, prob approaches 1.

Thm (CLT): If $X_1, \ldots X_n$ are iid samples from some distribution with $\mu, \sigma^2$ $\Rightarrow$ limiting dist of $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ (Std normal)

as $n \to \infty$ we express $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$

- If large sample taken from any dist. with $\mu, \sigma^2$
  - Std.ized mean $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$ approximately
  - Sample mean $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ approximately
  - Sum $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$ approximately

ex) CLT: $E(\bar X)=\mu$, $Var(\bar X)=\sigma^2$, $Sd(\bar X)=\sigma$

Assuming iid $\Rightarrow$ use CLT. $n=36>30$

1) $\mu_{\bar x}=200$, $\sigma_{\bar x}=\frac{\sigma}{\sqrt n}=2.5$

2) $P(\bar X\ge 204)=P(Z\ge 1.6)\simeq 1-\Phi(1.6)=0.0548$

ex) find dist $i=1,\dots,900$, $x_i=1$ or $x_i=0$ $\Rightarrow x_1,\dots x_{900}$

$E(x_i)=\frac12$, $var(x_i)=\frac14 \Rightarrow \mu=\frac12$, $\sigma^2=\frac14$

1) H is a normal r.v. $H=\sum_{i=1}^n X_i \Rightarrow n\bar X_n \overset{approx}{\sim} N(n\mu, n\sigma^2)$ by CLT

2) $var(H)=var(n\bar X_n)=n^2var(\bar X_n)=n^2\frac{\sigma^2}{n}=\frac{\sigma^2}{n}$

$E(H)=np=450$ $var(H)=np(1-p)=225$

$X_n\sim Bernoulli(p=\frac12)$, $P(\frac{H-450}{15}\ge\frac{495-450}{15})=P(Z\ge 3)=1-\Phi(3)=0.0013$

Case #1: $P(a\le\bar X\le b)=P(\frac{a-\mu}{\sigma_{\bar x}}\le z\le\frac{b-\mu}{\sigma_{\bar x}})\overset{CLT}{\approx}\Phi(\frac{b-\mu}{\sigma/\sqrt n})-\Phi(\frac{a-\mu}{\sigma/\sqrt n})$

Case #2: $x_1,\dots,x_n\sim Bernoulli(\theta)$, $\bar X=\frac1n\sum_{i=1}^n x_i$, $H=\sum_{i=1}^n x_i$ = # of success

$y=g(\bar x)$ find $g^{-1}$, $P(a\le y\le b)=P(g^{-1}(a)\le\bar x\le g^{-1}(b))$

Chi-square dist.

Thm 1: $X\sim N(0,1)\Rightarrow X^2\sim\chi^2_{\nu=1}$

Thm 2: $x_1,\dots,x_n \overset{iid}{\sim} N(0,1)\Rightarrow y=\sum_{i=1}^n x_i^2\sim\chi^2_{\nu=n}$ degrees of freedom $n$,

Thm 3: $x_1,\dots x_n$ independent r.vs having chi-squared with $\nu_1,\dots\nu_n$ df

$\Rightarrow y=\sum_{i=1} x_i\sim\chi^2_{\nu=\nu_1+\dots+\nu_n}$

$\bar X, s^2$ properties:

Thm 4: $X_1, X_2$ independent $\sim\chi^2_{\nu_1}$, df$=\nu$; and $X_1+X_2$ have chi square dist w/ $\nu>\nu_1\Rightarrow X_2$ has chi squared dist w/ $\nu_2=\nu-\nu_1$ df

Thm 5: $X_1,\dots X_n \overset{iid}{\sim} N(\mu,\sigma^2)$, $\bar X, S^2$ sample mean, variance then

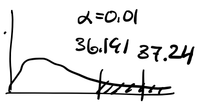
1) $\bar X, S^2$ independent

2) r.v. $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{n-1}$ chisquared with $n-1$ df.

chisquared dist: $(1-\alpha)100\%$ quantile $\chi^2_{\alpha,\nu}$: $P(X\ge\chi^2_{\alpha,\nu})=\alpha$

- $\nu\le 30$ values of $\chi^2_{\alpha,\nu}$ found on table.
- $\nu>30$ values approximated by normal dist.

$\nu\ge 30\Rightarrow\chi^2_\nu=\sum_{i=1}^{\nu} y_i$ : $y_i \overset{iid}{\sim}\chi^2_1$ apply CLT




ex) chisquared: $n=20$, $\sigma=0.06$ $\frac{(n-1)s^2}{\sigma^2}>\chi^2_{\alpha=.01,\nu=19}=36.191$ (given)

$\Rightarrow\frac{(20-1)(0.084)^2}{(0.06)^2}=37.24$, $P(s^2\ge S^2)\le 0.01\Rightarrow P(\frac{(n-1)s^2}{\sigma^2}\ge\frac{(n-1)S^2}{\sigma^2})\le 0.01$

Unknown $\sigma$: random sample from normal dist w/ $\mu,\sigma^2$

Std.ized mean $\frac{\bar X-\mu}{\sigma/\sqrt n}\sim N(0,1)$

- when $\sigma^2$ is unknown, we replace $S^2$, and define $T=\frac{\bar X-\mu}{S/\sqrt n}$
- $\bar X, S^2$ independent given $\sigma$, $\frac{\bar X-\mu}{\sigma/\sqrt n}\sim N(0,1)$ and $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{n-1}$ and define dist of T

T-distribution: $Y, Z$ independent r.vs, $Y\sim\chi^2_\nu$, $Z\sim N(0,1)$

$\Rightarrow T=\frac{Z}{\sqrt{Y/\nu}}$ given by $f(t)=\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$ : $x\in(-\infty,\infty)$

Thm 6: $X_i\overset{iid}{\sim}N(\mu,\sigma^2)\Rightarrow$ std.ized mean $T=\frac{\bar X-\mu}{S/\sqrt n}\sim t_{n-1}$

- $n\ge 30$ $\overset{approx}{\sim} N(0,1)$

t table: $(1-\alpha)100\%$ quantiles for $t_{\alpha,\nu}$ for $\alpha,\nu$ values. $P(X\ge t_{\alpha,\nu})=\alpha$

- $t_{1-\alpha,\nu}=-t_{\alpha,\nu}$ symmetric about $t=0$

ex) $n=16$, $\mu=12$, $\bar x=16.4$, $s=2.1\Rightarrow t=\frac{\bar x-\mu}{s/\sqrt n}=8.38$

F distribution: two r.vs $U\sim\chi^2_{\nu_1}$, $V\sim\chi^2_{\nu_2}$ are independent

$\Rightarrow F=\frac{U/\nu_1}{V/\nu_2}\sim F_{\nu_1,\nu_2}$ (not symmetric)

Thm 7: $S_1^2, S_2^2$ variances of independent random samples of size $n_1, n_2$ from normal populations w/ variances $\sigma_1^2,\sigma_2^2\Rightarrow$ variance ratio test

$$F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F_{n_1-1,n_2-1}$$

Order statistics: arrange $X_1,\dots X_n$ by size. $Y_1=\min(X_1,\dots X_n),\dots,Y_n=\max()$

Y is called the order statistic, denote $Y_r$ as the rth order statistic

ex) $n=2$, Y order statistic of X if $y_1=x_1, y_2=x_2$ : $x_1<x_2$

$y_1=x_2, y_2=x_1$ : $x_2<x_1$

…ding order statistic w/out norm assumption provides information on characteristics of dist

another def: arrange $X_1,\dots X_n$ by size, $X_1\dots X_n\overset{iid}{\sim}f$ and $y_1\le y_2\le\dots\le y_n$

$y_1=\min(x_1,\dots,x_n),\dots,y_n=\max(x_1,\dots x_n)$

distribution of $Y=Y_1,\dots Y_r: r=1,\dots,n$ ③ $f_{y_1}\leftarrow$ ② $F_{y_1}\leftarrow$ ① $1-F_{y_1}$

ex) cdf $F_{y_1}(y_1)=P(Y_1\le y_1)=1-\bar F_{y_1}(y_1)=1-P(Y_1\le y_1)=P(Y_1>y_1)$

$\Rightarrow P(Y_1>y_1)=P(X_1>y_1,\dots,X_n>y_1)=\prod_{i=1}^n P(X_i>y_1)=[P(X_1>y_1)]^n$

$\Rightarrow F_{y_1}(y_1)=1-[P(X_1>y_1)]^n=1-\left[\int_{y_1}^\infty f(x)dx\right]^n$

$\Rightarrow f_{y_1}(y_1)=\frac{d(F_{y_1}(y_1))}{dy_1}$

Thm: The random sample of size n from infinite population has value of $f(x)$ at $x$, the pdf of nth order statistic $y_r$:

$$f_r(y_r)=\frac{n!}{(r-1)!(n-r)!}\left[\int_{-\infty}^{y_r}f(x)dx\right]^{r-1}f(x)\left[\int_{y_r}^{\infty}f(x)dx\right]^{n-r} : y\in(-\infty,\infty)$$

ex) random sample of size n, from exponential population w/ parameter $\theta$, sampling dist of $y_1, y_n$?

pdf: $f(x)=\frac1\theta e^{-\frac{x}{\theta}}$: $x>0$, cdf: $F(t)=e^{-t/\theta}$: $t>0$, apply formula $f_1(y_1), f_n(y_n)$

$F(t)=\int_{-\infty}^t f(t)dt=e^{-t/\theta}$, $f_1(y_1)=nf(y_1)\left[\int_{y_1}^\infty f(x)dx\right]^{n-1}=n\frac1\theta e^{-(\frac{y_1}{\theta})}=\left[1-e^{-(y_1/\theta)}\right]^{n-1}$

$f_n(y_n)=nf(y_n)\left[\int_{-\infty}^{y_n}f(x)dx\right]^{n-1}=n\frac1\theta e^{(-y_n/\theta)}\left[e^{(\frac{-y_n}{\theta})}\right]=\frac n\theta e^{(\frac{-ny_n}{\theta})}$

**Review:**

#1) let $X_1,\dots X_n$ constitute a random sample from population: $f(x)=e^{-(x-\delta)}$: $x>\delta$ find the smallest sample value.

Ans: support of $y_1$: $y_1>\delta$, find pdf of $y_1\Rightarrow f_1(y_1)=nf(y_1)\left[\int_{y_1}^\infty f(x)dx\right]^{n-1}=ne^{-(x-\delta)}\left[\int_{y_1}^\infty e^{-(x-\delta)}dx\right]^{n-1}$ ①

① $\Rightarrow\int_{y_1}^\infty e^{-(x-\delta)}dx=\int_{y_1-\delta}^\infty e^{-x}dx=1-e^{-(y_1-\delta)}\Rightarrow ne^{-(y_1-\delta)}\left[1-e^{-(y_1-\delta)}\right]^{n-1}$ : $y_1>\delta$

#2) claim that var from normal pop is $\sigma^2=4$. rejected if variance of random sample $n=9$, $>7.7535$ find prob of rejecting this claim $\sigma^2=4$?

Ans: use $\chi^2$ dist: $S^2$, $n=9\Rightarrow\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{9-1}=\chi^2_8$ $P(\text{reject})=P(S^2\ge 7.7535)=P(\frac{(n-1)S^2}{\sigma^2}\ge\frac84(7.7..))$

$\Rightarrow P(\chi^2\ge\chi^2_8=15.507)=0.05$

#3) random sample $n=12$, from normal pop has $\bar x=27.8$ $s^2=3.24$ want to test pop $\mu$.

a) choose statistic, and find samp. dist. Ans: know $n,\bar x,S^2$ infer about $\mu\Rightarrow T=\frac{\bar X-\mu}{S/\sqrt n}\sim t_{n-1}$

b) claim $\mu=25$, surpasses 99%-quantile Ans: $\frac{\bar X-\mu}{S/\sqrt n}=\frac{27.8-25}{3.24/\sqrt{12}}\approx 5.3886>t_{12-1,0.01}=2.718$

#4) If $X_1\dots X_n$ constitute a random sample (size n) std dev=3, a) use chebyshev's to determine the smallest n samples that must be taken

Ans: $P(|\bar X_n-\mu|<\varepsilon)\ge 1-\frac{var(\bar X_n)}{\varepsilon^2}=1-\frac{\sigma^2}{n\varepsilon^2}$ (find n)

$\Rightarrow P(|\bar X_n-\mu|<.03)\ge 1-\frac{var(\bar X_n)}{(0.3)^2}\ge 1-\frac{9}{n(.9)}=1-\frac{100}{n}\ge 0.95\Rightarrow\begin{cases}1-\frac{100}{n}\ge 0.95\\ \frac{100}{n}\le 0.05\\ n\ge\frac{100}{0.05}=2000\end{cases}$

b) use CLT to determine smallest sample n, Ans: $P(|\bar X_n-\mu|\le 0.3)=P(|\frac{\bar X-\mu}{\sigma/\sqrt n}|\le\frac{0.3}{\sigma/\sqrt n})$

$\Rightarrow\overset{CLT}{\approx}P(Z\le\frac{0.3\sqrt n}{3})=.95\Rightarrow\frac{\sqrt n}{10}=Z_{0.95}=2.57\Rightarrow n=(25.7)^2$

#3) random sample of $n=225$ taken from exponential pop. w/ $\theta=4$

Based on CLT what's the prob mean of sample will exceed 4.5?

Ans: let $(X_1,\dots X_{225})=X\sim Exp(\theta=4)\Rightarrow f(x)=\frac14 e^{-\frac x4}$: $x>0$, $\Rightarrow E(X)=4$, $Var(X)=4^2$

$\Rightarrow P(\bar X\ge 4.5)=P(\frac{\bar X-\mu_{\bar x}}{\sigma_{\bar x}}>\frac{4.5-4}{\sqrt{4^2/225}})\overset{CLT}{\approx}P(Z>1.875)=0.0304$

#4) proportion of families own house is 0.70. If 84 families surveyed $\sim Bernoulli(\theta=0.7)$ w/ what prob. can we assert that value obtained from sample proportion $\hat\theta$ falls between $(.64,.76)$

Ans: $\hat\theta=\frac1n\sum x_i$, we have $\mu_{\hat\theta}=.7$, $\sigma^2_{\hat\theta}=\frac{\sigma^2}{n}=\frac{\theta(1-\theta)}{n}=\frac{(.7)(.3)}{84}=0.0025$

a) Chebyshev's: $P(.64<\hat\theta<.76)=P(.64-.7<\hat\theta-\mu_{\hat\theta}<.76-.7)$

$\Rightarrow P(-.06<\hat\theta-\mu_{\hat\theta}<.06)=P(|\hat\theta-\mu_{\hat\theta}|<0.06)\ge 1-\frac{\sigma^2_{\hat\theta}}{.06^2}=1-\frac{0.0025}{(0.06)^2}=.3056$

b) CLT: $P(.64<\hat\theta<.76)=P(|\hat\theta-\mu_{\hat\theta}|<0.06)=P(|\frac{\hat\theta-\mu_{\hat\theta}}{\sigma_{\hat\theta}}|<\frac{0.06}{\sigma_{\hat\theta}})=P(|Z^*|<1.2)=2\cdot(\Phi)$

$=2(\Phi(1.2))=.7698$

#5) Integrate appropriate chi-squared density to find the prob. that the variance of a random sample of size $n=5$ from normal population w/ $\sigma^2=25$ falls between 20 and 30?

Ans: know $n=5$, $\sigma^2=25$, $\Rightarrow P(20<S^2<30)=P(\frac{(n-1)20}{\sigma^2}<\frac{(n-1)S^2}{\sigma^2}<\frac{(n-1)30}{\sigma^2})=P(3.2<\chi^2_{5-1}<4.8)$

$=\frac{1}{2^{4/2}\Gamma(\frac42)}\int_{3.2}^{4.8}x^{4/2-1}e^{-\frac x2}dx=\frac{1}{4(2-1)!}\int_{3.2}^{4.8}xe^{-\frac x2}dx=\begin{cases}u=x & v=-2e^{-\frac x2}\\ du=dx & dv=e^{-\frac x2}dx\end{cases}$

$=\frac14\left(-2xe^{-\frac x2}\Big|_{3.2}^{4.8}+2\int_{3.2}^{4.8}e^{-\frac u2}du\right)=-3.4e^{-2.4}+2.6e^{-1.6}\approx 0.216$

ex) $n_1=30$, $n_2=50$, $\mu_1=78$, $\mu_2=75$, $\sigma_1^2=150$, $\sigma_2^2=200$ Find prob. that mean of first sample will exceed 2nd by at least 4.8 Ans: $Y=\bar X_1-\bar X_2$, $\mu_y=\mu_1-\mu_2=3$, $\sigma_y^2=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}=\frac{150}{30}+\frac{200}{50}=9$

$\Rightarrow P(Y\ge 4.8)=P(\frac{Y-\mu_y}{\sigma_y}\ge\frac{4.8-\mu_y}{\sigma_y})=P(Z^*\ge\frac{4.8-3}{3})\Rightarrow 0.5-\Phi(.6)=0.2743$

ex) $S_1^2, S_2^2$ are variances of independent random samples of size $n_1=10$, $n_2=15$ from normal populations w/ equal variances find $P(\frac{S_1^2}{S_2^2}<4.03)$ Ans: $P_{\sigma_1=\sigma_2}(\frac{S_1^2}{S_2^2}<4.03)=P(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}<\frac{4.03\sigma_2^2}{\sigma_1^2})$

$\Rightarrow P(\frac{\chi^2_{n_1-1}/n_1-1}{\chi^2_{n_2-1}/n_2-1}<4.03)=P(F_{n_1-1,n_2-1}<4.03)=1-P(F_{9,14}\ge 4.03)=1-.01=.99$

| | | | |
|---|---|---|---|
| $Z=\frac{\bar X-\mu}{\sigma/\sqrt n}$ | know $\sigma^2,\mu,n$ | $\Rightarrow Z\sim N(0,1)$ | infer $\mu$ use $\bar X$ |
| $T=\frac{\bar X-\mu}{S/\sqrt n}$ | know $S^2,\mu,n$ | $\Rightarrow T\sim t_{n-1}$ | infer $\mu$, use $\bar X$ |
| $\chi^2=\frac{(n-1)S^2}{\sigma^2}$ | know $S^2,\sigma^2$ | $\Rightarrow\chi^2\sim\chi^2_{n-1}$ | infer $\sigma^2$ use $S^2$ |
| $F=\frac{(S_1/\sigma_1^2)}{(S_2/\sigma_2^2)}$ | know $S_1^2,S_2^2,\sigma_1^2,\sigma_2^2$ | $\Rightarrow F\sim F_{S_1,S_2}$ | infer ratio $\frac{\sigma_1^2}{\sigma_2^2}$ |

Finite population: joint pdf of finite population (size $N$) is $f(x_1, \ldots x_n) = \frac{1}{N(N-1)\ldots(N-n+1)}$.

marginal dist of $X_r$: $f(x_r) = \frac{1}{N}$ : $x_r = c_1, \ldots c_N$ for $r = 1, \ldots, n$ (discrete Unif)

for $\{c_1, \ldots c_N\}$ $\mu = \sum_{i=1}^{N} c_i \frac{1}{N}$, $\sigma^2 = \sum_{i=1}^{N} (c_i - \mu)^2 \frac{1}{N}$

Joint marginal dist of any two r.v.'s from $(x_1, \ldots x_n)$ $(x_r, x_s)$

$g(x_r, x_s) = \frac{1}{N(N-1)}$ : $(x_r, x_s) \in \{c_1, \ldots c_N\}$ $x_r \neq x_s$

Covariance $(x_r, x_s) = \mathrm{Cov}(x_r, x_s) = \frac{-\sigma^2}{N-1}$

Thm: $\bar{X}$ mean of random sample taken without replacement from finite population size $N$, with $\mu, \sigma^2$

$\Rightarrow E(\bar{X}) = \mu$, $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$ Correction factor $\leq 1$