MVE055-EXAM-NOTES

Lukas Rahmn

Probability

$$\binom{n}{r}=\frac{n!}{r!(n-r)!}$$

$$P[A_1\cup A_2]=P[A_1]+P[A_2]-P[A_1\cap A_2]$$

$$P[A_2|A_2]=\frac{P[A_1\cap A_2]}{P[A_1]}$$

$$P[A_j|B]=\frac{P[B|A_j]P[A_j]}{\sum_{i=1}^{n}P[B|A_i]P[A_i]}$$

Series

$$\sum_{k=1}^{\infty}ar^{k-1}=\frac{a}{1-r}\ \ |r|<1$$

$$\sum_{k=1}^{n}ar^{k-1}=\frac{a(1-r^n)}{1-r}$$

Arithmetic Series: $S_n=\frac{n(a_1+a_n)}{2}$

Expectancy & Variance

$$E[H(X)]=\sum_{x\in X}H(x)f(f)$$

$$E[H(X)]=\int_{-\infty}^{\infty}H(x)f(x)dx$$

$E[c]=c,\ E[cX]=cE[X]$

$E[X+Y]=E[X]+E[Y]$

$E[XY]=E[X]E[Y]$ *iff indepentent*

$\sigma^2=Var[X]=E[(X-\mu)^2]=E[X^2]-E[X]^2$

$Var[c]=0,\ Var[cX]=c^2Var[X]$

$Var[aX+bY]=a^2Var[X]+b^2Var[Y]+abCov(X,Y)$

$Cov(X,Y)=E[(X-\mu_x)(Y-\mu_y)]$

$Cov(X,Y)=E[XY]-E[X]E[Y]$

Moment generating functions

$m_X(t)=E[e^{tX}]$

$$\left.\frac{d^k m_X(t)}{dt^k}\right|_{t=0}=E[X^k]$$

$Y=X_1+\ldots X_n\Longleftrightarrow m_y(t)=m_{x_1}(t)*\ldots m_{x_n}(t)$

$Y=\alpha+\beta X\rightarrow m_Y(t)=e^{\alpha t}m_X(\beta t)$

Discrete distributions

$X\sim Geo(p)\rightarrow F_x(t)=1-(1-p)^{\text{floor}(t)}$

$$P[X>x]=1-P[X\le x]=1-F_X(x)=\sum_{x\in X}f(x)$$

$f_X(t)=F_X(t)-F_X(t-1)$

Continuous distributions

$$\int_{-\infty}^{\infty}f_X(x)dx=1$$

$$P[a\le X\le b]=\int_a^b f_X(x)dx$$

$P[X=x]=0$

$P[X\ge x]=1-P[X\le x]=1-F_X(x)$

$$F_X(t)=\int_{-\infty}^{t}f_X(x)dx$$

$$f_X(x)=\frac{dF_X(x)}{dx}$$

Continuity correction

Discrete	Continuous
$P[X=x]$	$P[x-0.5<X<x+0.5]$
$P[X\le x]$	$P[X<x+0.5]$
$P[X<x]$	$P[X<x-0.5]$
$P[X\ge x]$	$P[X<x-0.5]$
$P[X>x]$	$P[X<x+0.5]$

Chebyshev's inequality

$$P[|X-\mu|\ge\epsilon]\le\frac{\sigma^2}{\epsilon^2}$$

$$P[|X-\mu|<\epsilon]>1-\frac{\sigma^2}{\epsilon^2}$$

$$\epsilon=k\sigma\rightarrow P[|X-\mu|<k\sigma]\le 1-\frac{1}{k^2}$$

Joint Distributions

$$f_{XY}=\sum_{x\in X}\sum_{y\in Y}f_{XY}(x,y)=1$$

$$f_X(x)=\sum_{y\in Y}f_{XY}(x,y)$$

$$f_Y(y)=\sum_{x\in X}f_{XY}(x,y)$$

$$f_{XY}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{XY}(x,y)dxdy=1$$

$$f_X(x)=\int_{-\infty}^{\infty}f_{XY}(x,y)dy$$

$$f_Y(y)=\int_{-\infty}^{\infty}f_{XY}(x,y)dx$$

$f_{XY}(x,y)=f_X(x)f_Y(y)$ iff indepentent

$$E[H(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}H(x,y)f(x,y)dxdy$$

$E[XY]=E[H(X,Y)],H(X,Y)=X*Y$

$$E[X]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{XY}(x,y)dxdy$$

$$E[Y]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf_{XY}(x,y)dxdy$$

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}}$$

Sample statistics

$$\bar{X}=\sum_{i=1}^{n}\frac{X_i}{n}$$

Median: $P[X\le m]=P[X\ge m]=\frac{1}{2}$

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

$$=\frac{n\sum_{i=1}^{n}X_i^2-\left(\sum_{i=1}^{n}X_i\right)^2}{n(n-1))}$$

$\sigma\sim$(estimated range)/4, normal

$\sigma\sim$(estimated range)/6, unkown

Unbiased: $E[\hat{\theta}]=\theta$

$$Var[\bar{X}]=\frac{\sigma^2}{2}$$

Confidence interval

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

$\sim chi-squared(n-1)$

$$\frac{Z}{\sqrt{\chi^2_\gamma/\gamma}}\sim T_\gamma,\ \frac{\bar{X}-\mu}{\sqrt{S^2/n}}\sim T_{(n-1)}$$

$$\sigma^2\in\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}},\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right]$$

General: find $P[-D_{(\alpha/2)}\le C\le D_{(\alpha/2)}]$

σ^2 known, $\bar{x}\pm Z_{(\alpha/2)}\sigma\sqrt{n},Z\sim N(0,1)$

σ^2 unknown, $\bar{x}\pm t_{\alpha/2}S/\sqrt{n}$

$$n=\frac{(Z_{\alpha/2})^2\sigma^2}{d^2},n=\frac{(Z_{\alpha/2})^2\hat{\sigma}^2}{d^2}$$

Central limit theorem

For a sample of size n from distribution with mean μ Variance σ^2, when n is large $\bar{X}$ is approximately normal with mean μ Variance σ^2/n.

Hypothesis testing

$\alpha=P[$"type 1 error"$]=P[$"reject H_0"$|H_0]$

$\beta=P[$"type 2 error"$]=P[$"fail reject H_0"$|H_0^c]$

"Power of test"$=1-\beta$

Non parametric methods

Sign-test: Calculate median M on the sample of n. Then calculate $X_i-M$for$x_i\in X$. Let Q_+ denote the number of positive differences. $Q_+\sim Bin(n,\frac{1}{2})$. Zeros are assigned two the sign that supports H_0

Wilcoxon rank-test: As befor calculate M and X_i-M order the list of differences from least to greatest. Number them 1 to N where 1 is the smalest difference. For equal difference assign the mean rank to each of them. Let W_+ denote the sum of positive ranks and W_- the negative ranks. $W=min(W_+,W_-)$ Lookup w in table or use normal approximation. This test can also be used for paired data, then use X_i-Y_i instead of X_i-M_x

$$E[W]=\frac{n(n+1)}{4},Var[W]=\frac{n(n+1)(2n+1)}{24}$$

$$\alpha=W\le c\text{ or }W\ge C,C=\frac{n(n+1)}{2}-c$$

$$W=\sum_{j=1}^{n}R_jI_j,\text{max }W\rightarrow I_j=1.$$

Wilcoxon rank-sum test: Given two samples $X_1,X_2,...,X_n$ and $y_1,y_2,...,y_m,\ m\le n$ pool them and order them from smalest to largest. Assign ranks from 1 to N+M, draws are each assigned the average rank. Let W_m= the sum of all ranks beloning to y, the smaller sample. Let c denote the lower critical boundry left tailed tests and C the upper boundry. c is given i given in the table.

$$C=m(m+n+1)-c$$

$$E[W_m]=\frac{m(m+n+1)}{2}$$

$$Var[W_m]=\frac{mn(m+n+1)}{12}$$

Propotions

$\hat{p}=\frac{X}{n},\ \hat{p}\pm Z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$

$$n=\frac{Z^2_{\alpha/2}\hat{p}(1-\hat{p})}{d^2},\hat{p}\text{ estimate known}$$

$$n=\frac{Z^2_{\alpha/2}}{4d^2},\ \text{ estimate unknown}$$

$$T=\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$$

$$\widehat{p_1-p_2}=\frac{X_1}{n_1}-\frac{X_2}{n_2}$$

$$\widehat{p_1-p_2}\pm Z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1}+\frac{\hat{p_2}(1-\hat{p_2})}{n_2}}$$

When: $p_1=p_2,\hat{p_p}=\frac{n_1p_1+n_2p_2}{n_1+n_2}$

$$T=\frac{\hat{p_1}-\hat{p_2}}{\hat{p_p}(1-\hat{p_p})(1/n_1+1/n_2)}$$

$$n=Z^2_{\alpha/2}\frac{\hat{p_1}(1-\hat{p_2})+\hat{p_2}(1-\hat{p_2})}{d^2}$$

$$n=\frac{Z_{\alpha/2}^2}{2d^2}$$

Comparing means

$$\widehat{\mu_1-\mu_2}=\bar{X}_1+\bar{X}_2\sim N(\mu_1-\mu_2,\sigma_1^2+\sigma_2^2)$$

$$\frac{\chi^2_{\gamma_1}/\gamma_1}{\chi^2_{\gamma_2}/\gamma_2}\sim F_{\gamma_1,\gamma_2},\ \ \frac{S_1^2}{S_2^2}\sim F_{n_1-1,n_2-1}$$

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

Comparing means:

Variance known:

$$T=\frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{(\sigma_1^2/n_1+\sigma_2^2/n_2)}}\sim N(\mu_1-\mu_2,\sigma_1^2+\sigma_2^2)$$

Variance unknown:

$$\sigma_1^2=\sigma_2^2, T=\frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{S_p^2(1/n_1+1/n_2)}}\sim T_{n_1+n_2-2}$$

$$\sigma_1^2\neq\sigma_2^2, T=\frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{(S_1^2/n_1+S_2^2/n_2)}}\sim T_\gamma$$

$$\gamma=\frac{(S_1^2/n_1+S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1-1}+\frac{(S_2^2/n_2)^2}{n_2-1}}$$

Linear regression

$$\mu_{Y|x}=\beta_0+\beta_1 x,\ \ Y|x_i=\beta_0+\beta_1 x_i+E_i$$

$$E_i\sim N(0,\sigma^2), Y|x\sim N(\mu_{Y|x},\sigma^2)$$

$$\text{SSE}=\sum_{i=1}^n e_i^2=\sum_{i=1}^n (y_i-b_0-b_1x_i)^2$$

$$\text{Minimize SEE:}\frac{dSEE}{db_0}=-2\sum_{i=1}^n(y_i-b_0-b_1x_i)$$

$$\frac{dSEE}{db_1}=-2\sum_{i=1}^n(y_i-b_0-b_1x_i)x_i$$

$$b_1=\frac{n\sum_{i=1}^n x_iy_i-\left(\sum_{x=i}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n\sum_{i=1}^n x_i^2-\left(\sum_{i=1}^n x_i\right)^2}$$

$$b_0=\bar{y}-b_1\bar{x}, Y_i\sim N(\beta_0+\beta_1x_i,\sigma^2)$$

$$S_{xx}=\sum(x_i-\bar{x})^2,\ \ S_{yy}=\sum(Y_i-\bar{Y})^2$$

$$S_{xx}=\frac{n\sum x_i^2-\left(\sum x_i\right)^2}{n}=\sum x_i^2-n\bar{x}^2$$

$$S_{xy}=(x_i-\bar{x})(Y_i-\bar{Y})=\sum(x_i-\bar{x})Y_i$$

$$S_{xy}=\frac{(n\sum x_iY_i-\sum x_i\sum Y_i)}{n}$$

$$b_1=\frac{S_{xy}}{S_{xx}}=\frac{\sum(x_i-\bar{x})Y_i}{S_{xx}}=c_1Y_1+\cdots+c_nY_n$$

$$c_j=\frac{(x_j-\bar{x})Y_j}{S_{xx}},\ \ b_1\sim N\left(\beta_1,\frac{\sigma^2}{S_{xx}}\right)$$

$$b_0\sim N\left(\beta_0,\frac{\sum x_i^2}{nS_{xx}}\sigma^2\right)$$

$$S^2=\hat{\sigma}^2=\frac{SEE}{n-2}=S_{yy}-S_{xx}b_1^2$$

$$T_{n-2}=\frac{b_1-\beta_1^0}{S/\sqrt{S_{xx}}}, b_1\pm t_{\alpha/2}S/\sqrt{S_{xx}}$$

$$T_{n-2}=\frac{b_0-\beta_0^0}{\left(\frac{S\sqrt{\sum x^2}}{\sqrt{nS_{xx}}}\right)}, b_0\pm t_{\alpha/2}S\sqrt{\frac{\sum x^2}{nS_{xx}}}$$

$$\hat{\mu}_{Y|x}\sim N\left(\mu,\sigma^2\left[\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}\right]\right)$$

$$\hat{\mu}_{Y|x}\pm t_{\alpha/2}S\sqrt{\frac{1}{n}+\frac{(x-\bar{x}^2)}{S_{xx}}}$$

$$(\hat{Y}|x-Y|x)=(\hat{\mu}_{Y|x}-Y|x)$$

$$(\hat{Y}|x-Y|x)\sim N\left(0,\sigma^2\left[1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}\right]\right)$$

$$T_{n-2}=\frac{\hat{Y}|x-Y|x}{S\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}}$$

$$\hat{Y}|x\pm t_{\alpha/2}S\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x}^2)}{S_{xx}}}$$

Partial integration

$$\int u(x)v'(x)dx=u(x)v(x)-\int u'(x)v(x)dx$$

Name	density	def	mgf	μ	σ^2
Gemetric	$(1-p)^{x-1}p$	$x\geq 1$	$\frac{pe^t}{1-qe^t}$	p^-1	qp^{-2}
Uniform	$\frac{1}{n}$	$x=x_1\dots x_n$	$\frac{\sum_{i=1}^n e^{tx}}{n}$	$\frac{\sum_{i=1}^n x_i}{n}$	$\frac{\sum_{i=1}^n x_i^2}{n}-\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2$
Binomial	$\binom{n}{x}p^x(1-p)^{n-x}$	$x=0,1\dots,n$	$(q+pe^t)^n$	np	$np(1-p)$
Poisson	$\frac{e^{-k}k^x}{x!}$	$x=0,1,2,\dots$	$e^{k(e^x-1)}$	k	k
Exponential	$\frac{1}{\beta}e^{-x/\beta}$	$x>0,\beta>0$	$(1-\beta t)^{-1}$	β	β^2
Uniform	$\frac{1}{b-a}$	$a<x<b$	$\frac{e^{tb}-e^{ta}}{t(b-a)},\ t\neq 0,1,\ t=0$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	$\frac{1}{\sigma\sqrt{2\pi}}exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$	$-\infty<x<\infty$	$e^{\mu t+\sigma^2t^2/2}$	μ	σ^2
Chi-squared	$\frac{1}{\Gamma(\gamma/2)2^{\gamma/2}}x^{\gamma/2-1}e^{-x/2}$	$x>0$	$(1-2t)^{-\gamma/2}$	γ	2γ