REVIEW for FINAL EXAM

PROBABILITY Ch. 2

- $P(\phi) = 0$, $P(S) = 1$, $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

- BINOMIAL COEFF. $\binom{n}{y} = \dfrac{n!}{y!\,(n-y)!}$ # of combinations of n distinct items, take y at a time

Ex DECK of 52 playing cards. Select 1 card

~~P(B)~~ B_1 = card is a king ~~of spade~~ $P(B_1) = \dfrac{4}{52}$

B_2 = card is a heart $P(B_2) = \dfrac{13}{52}$

$P(B_1 \cap B_2) = \dfrac{1}{52}$

$P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2)$

Select 5 cards

B_1: all cards are hearts $P(B_1) = \dfrac{\binom{13}{5}}{\binom{52}{5}}$

B_2: at least one card is spade

Set B_2' = no cards are spade: $P(B_2') = \dfrac{\binom{39}{5}}{\binom{52}{5}}$

$P(B_2) = 1 - P(B_2')$

B_3: there are exactly 2 jacks, 2 queens, 1 ace

$$P(B_3) = \frac{\binom{4}{2}\binom{4}{2}\binom{4}{1}}{\binom{52}{5}}$$

- CONDITIONAL PROBABILITY $P(A|B) = \dfrac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B)\,P(B)$

Ex 3 cards are dealt at random from a 52 playing card deck.

B: all card are spade. A_1: one card is spade: $P(A_1) = \dfrac{13}{52}$

$$P(B) = P(A_1) \cdot P(A_2|A_1)\,P(A_3|A_1 \cap A_2) = \frac{13}{52} \cdot \frac{12}{51} \cdot \frac{11}{50}$$

$= \cancel{\frac{12}{52}\,\frac{12}{51}\,\frac{11}{50}}$

SAME as $P(B) = \dfrac{\binom{13}{3}}{\binom{52}{3}}$

• BAYES' THM   let A_1, A_2 be mutually exclusive and exhaustive events

$$P(A_1|B) = \frac{P(B|A_1)\,P(A_1)}{P(B|A_1)\,P(A_1) + P(B|A_2)\,P(A_2)}$$ (also with $A_1, A_2, \ldots$

Ex → Test 1, Ex 4

A disease affect 4% of population. A screening test is applied but test has 20% probability of failing to recognize disease. The Prob. of falsely concluding that disease is present is 8%.

let D: disease.   TP: test positive,   TN: test negative

$P(D) = 0.04$, $P(TN|D) = 0.20$, $P(TP|D') = 0.08$

What is Prob. of testing a person and finding that test is positive?

$$P(D|TP) = \frac{P(TP|D)\,P(D)}{P(TD|D)\,P(D) + P(TP|D')\,P(D')} = \frac{0.80 \cdot 0.04}{0.80 \cdot 0.04 + 0.08 \cdot (1-0.04}$$

$P(TP|D) = 1 - P(TN|D)$
$= 0.80$

DISCRETE R.V.

X r.v. $X = x_i,\ i=1,\ldots,n$ It varies over a range

$f(x)$ p.d.f → $\sum_{i=1}^{n} f(x_i) = 1$ → $f(x_i) = P(X = x_i)$

$$\mu = E(X) = \sum_{i=1}^{n} x_i f(x_i), \qquad \sigma^2 = E[(X-\mu)^2] = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2$$

Ex

x	1	2	3	4
f(x)	0.2	0.3	0.3	0.2

$\mu = 1 \cdot 0.2 + 2 \cdot 0.3 + 3 \cdot 0.3 + 4 \cdot 0.2$
$= 0.2 + 0.6 + 0.9 + 0.8 = 2.5$

$\sigma^2 = \sum_{i=1}^{4} x_i^2 f(x_i) - (2.5)^2$

BINOMIAL DISTRIBUTION

$Y = \sum_{i=1}^{n} X_i$ # of successes of n Bernoulli trials

pdf $f(y) = P(Y=y) = \binom{n}{y} p^y (1-p)^{n-y}$ $y = 0,1,\ldots,n$

BINOMIAL PDF $b(n,p)$

Note $E(Y) = np$, $\sigma^2 = np(1-p)$

Ex $P(\text{roof leak}) = p = 0.1$

Y = # of roofs leaky in 6 buildings $\sim b(n=6, p=0.1)$

$$P(Y \geq 2) = 1 - P(Y \leq 1) = 1 - \sum_{y=0}^{1} \binom{6}{y} (0.1)^y (0.9)^{6-y}$$

Ex In a pool 4 out 10 fish are tagged

$n=3$ fish are caught at random

$$P(2 \text{ or fewer are tagged}) = P(Y \leq 2) = \sum_{y=0}^{2} \binom{3}{y} p^y (1-p)^{n-y}$$

NEGATIVE BINOMIAL pdf

$$h(z) = \binom{z-1}{r-1} p^r (1-p)^{z-r} \qquad z = r, r+1, \ldots$$

Z = # of trials need to observe the r-th success

Ex A process continues until first defective item is produced

Probability of defects is $p = 0.05$

What is probability that the first defect occur at fifth trial?

$$P(Z=5) = \binom{4}{0} (0.05)(0.95)^4$$

$$P(Z \leq 5) = \sum_{z=1}^{5} \binom{z-1}{0} (0.05)(.95)^{z-1}$$

HYPERGEOMETRIC PDF

$N = N_1 + N_2$ objects. W = # items in first class out of n items

$$P(W=w) = \frac{\binom{N_1}{w}\binom{N_2}{w-n}}{\binom{N_1+N_2}{n}} \qquad w = 0, 1, \ldots n$$

POISSON PDF

Y : # of events/flaws happening over a period of time/space, assuming that λ, the average events/flaws per unit time/space is constant

$$p(y) = \frac{\lambda^y e^{-\lambda}}{y!} \qquad \mu = \lambda, \quad \sigma^2 = \lambda$$

Ex Flaws in drapery appear an average of 1 in 200 sq. feet. Use Poisson pdf

X = # flaws in 300 sq. feet

$$P(X \leq 1) = \sum_{y=0}^{1} \lambda^y \frac{e^{-\lambda}}{y!} \qquad \lambda = 1.5$$

Multivariate Discrete r.v.

$$f(x,y) = P(X=x, Y=y) \qquad (x,y) \in R$$

$$f_1(x) = \sum_y f(x,y), \qquad f_2(y) = \sum_x f(x,y) \qquad \text{MARGINAL PDF}$$

$$\mu_X = \sum_x x f_1(x), \qquad \mu_Y = \sum_y y f_2(y)$$

$$\operatorname{var}(X) = \sum_x (x-\mu_X)^2 f_1(x) \qquad \operatorname{var}(Y) = \sum_y (y-\mu_Y)^2 f_2(y)$$

$$\sigma_{XY} = \operatorname{cov}(X,Y) = \sum_{(x,y)\in R}\sum (x-\mu_X)(y-\mu_Y) f(x,y)$$

$$\rho = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}Y}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

X, Y are INDEP $\iff$ $f(x,y) = f_1(x) f_2(y)$

X, Y indep $\Rightarrow$ $\rho = 0$ $\qquad$ (But $\rho = 0 \not\Rightarrow X, Y$ indep.)

RK $\quad$ let $W = a_0 + a_1 X + a_2 Y$

$$\mu_W = a_0 + a_1 \mu_X + a_2 \mu_Y$$

$$\sigma^2_W = a_1^2 \sigma_X^2 + a_2^2 \sigma_Y^2 + 2a_1 a_2 \sigma_{XY}$$

Ch.3 CONTINUOUS R.V.

- X : continuous r.v.

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

Ex UNIFORM P.D.F. $f(x) = \frac{1}{b-a}$ $x \in [a,b]$

EXPECTATION $E[X] = \int x f(x)\,dx = \mu_X$

VARIANCE $var[X] = \int (x-\mu_X)^2 f(x)\,dx = \int x^2 f(x)\,dx - \mu_X^2$

Ex UNIFORM P.D.F. $\mu = \frac{b+a}{2}$, $\sigma^2 = \frac{(b-a)^2}{12}$

↳ See Prob. #1, Test #3

- Normal P.D.F. $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $x \in \mathbb{R}$

MEAN $= \mu$, VARIANCE $= \sigma^2$

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$ → use tables

↳ See ~~Prob~~ ex. 3.2-3 in Review Quiz #3

- Percentile Find x_p s.t. $P(X \le x_p) = p$

Ex $p = 0.95$ $P\left(\frac{X-\mu}{\sigma} \le \frac{x_p-\mu}{\sigma}\right) = 0.95$

$\Rightarrow P(Z \le z_p) = 0.95 \rightarrow z_p = 1.645$

$\frac{x_p-\mu}{\sigma} = 1.645 \Rightarrow \underline{x_p = 1.645 \cdot \sigma + \mu}$

- Other distributions Ex. Exponential Distribution

$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$ $x \ge 0$ $\mu = \beta$, $\sigma^2 = \beta^2$

- Distributions of several variables

Ch. 4 SAMPLING DISTRIBUTION

Sampling Mean $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ Suppose $E(X_i) = \mu$, $var(X_i) = \sigma^2$

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

CENTRAL LIMIT THM.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Prob. 2, Test #1

$\bar{X}$ sample mean from sample of size $n = 27$ taken from unif. distribution $f(x) = \frac{1}{6}$ $x \in [0,6]$

$\mu_X = 3$, $\sigma_X^2 = \frac{36}{12} = 3 \Rightarrow \mu_{\bar{X}} = 3$, $\sigma_{\bar{X}}^2 = 3/27 = \frac{1}{9}$

$$P(2.7 < \bar{X} < 3.2) = P\left(\frac{2.7-3}{1/3} < Z < \frac{3.2-3}{1/3}\right) = \Phi(0.6) - \Phi(-0.9)$$

CONFIDENCE INTERVAL for MEAN

- If (X_i) are NORMAL or n large + σ^2 known $\quad \bar{x} \pm z(\frac{\alpha}{2}) \frac{\sigma}{\sqrt{n}}$
- If (X_i) are NORMAL but σ^2 unknown $\quad \bar{x} \pm t(\frac{\alpha}{2}; n-1) \frac{\sigma}{\sqrt{n}}$
- If (X_i), (Y_i) NORMAL, σ_1^2, σ_2^2 known or n_1, n_2 large $\quad (\bar{x} - \bar{y}) \pm z(\frac{\alpha}{2}) \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
- If (X_i), (Y_i) NORMAL, σ_1^2, σ_2^2 unknown but equal $\quad (\bar{x} - \bar{y}) \pm t(\frac{\alpha}{2}; n_1+n_2-2) \sqrt{\frac{(n_1-1)S_x^2 + (n_2-1)S_y^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

CONFIDENCE of INTERVAL for PROPORTIONS

p binomial $b(n,p)$, n large $\quad \frac{Y}{n} \pm z(\frac{\alpha}{2}) \sqrt{\frac{Y/n\,(1 - Y/n)}{n}}$

$p_1 - p_2$ binomial, n_1, n_2 large $\quad \left(\frac{Y_1}{n_1} - \frac{Y_2}{n_2}\right) \pm z(\frac{\alpha}{2}) \sqrt{\frac{Y_1/n_1 (1 - Y_1/n_1)}{n_1} + \frac{Y_2/n_2 (1 - Y_2/n_2)}{n_2}}$

CONFIDENCE of INTERVAL for VARIANCES

$$\left[\frac{(n-1)s^2}{\chi^2(\frac{\alpha}{2}; n-1)}, \frac{(n-1)s^2}{\chi^2(1-\frac{\alpha}{2}; n-1)}\right]$$

TESTS of HYPOTHESIS

Use $z(\alpha)$ if σ is known, or $t(\alpha; n-1)$ otherwise

Tests of hypotheses

Hypotheses	Rejection Region of H_0
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$\dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \geq z(\alpha)$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$\dfrac{\overline{x} - \mu_0}{s/\sqrt{n}} \geq t(\alpha; n - 1)$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$\dfrac{\overline{x} - \overline{y}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \geq z(\alpha)$
$H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	$\dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \geq t(\alpha; n_1 + n_2 - 2)$
$H_0: p = p_0$ $H_1: p > p_0$	$\dfrac{y/n - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} \geq z(\alpha)$
$H_0: p_1 = p_2$ $H_1: p_1 > p_2$	$\dfrac{y_1/n_1 - y_2/n_2}{\sqrt{\left(\dfrac{y_1 + y_2}{n_1 + n_2}\right)\left(1 - \dfrac{y_1 + y_2}{n_1 + n_2}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \geq z(\alpha)$
$H_0: \beta_1 = \beta_{10}$ $H_1: \beta_1 > \beta_{10}$	$\dfrac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\dfrac{\sum (y_i - \hat{y}_i)^2/(n - 2)}{\sum (x_i - \overline{x})^2}}} \geq t(\alpha; n - 2)$

Confidence intervals

Parameter	Assumptions	Endpoints
μ	$N(\mu, \sigma^2)$ or n large σ^2 known	$\bar{x} \pm z(\alpha/2)\dfrac{\sigma}{\sqrt{n}}$
μ	$N(\mu, \sigma^2)$ σ^2 unknown	$\bar{x} \pm t(\alpha/2; n-1)\dfrac{s}{\sqrt{n}}$
$\mu_1 - \mu_2$	Independent Distributions σ_1^2, σ_2^2 known n_1, n_2 large	$\bar{x} - \bar{y} \pm z(\alpha/2)\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$
$\mu_1 - \mu_2$	Independent Normal Distributions σ_1^2, σ_2^2 unknown but equal	$\bar{x} - \bar{y} \pm t(\alpha/2; n_1 + n_2 - 2)$ $\times \sqrt{\dfrac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$
p	Binomial $b(n, p)$ n large	$\dfrac{y}{n} \pm z(\alpha/2)\sqrt{\dfrac{(y/n)(1 - y/n)}{n}}$
$p_1 - p_2$	Independent Binomial Distributions n_1, n_2 large	$\dfrac{y_1}{n_1} - \dfrac{y_2}{n_2} \pm z(\alpha/2)$ $\times \sqrt{\dfrac{(y_1/n_1)(1 - y_1/n_1)}{n_1} + \dfrac{(y_2/n_2)(1 - y_2/n_2)}{n_2}}$
β_1	$N(\beta_0 + \beta_1 x, \sigma^2)$	$\hat{\beta}_1 \pm t(\alpha/2; n-2)\sqrt{\dfrac{\Sigma(y_i - \hat{y}_i)^2/(n-2)}{\Sigma(x_i - \bar{x})^2}}$