STAT 324 Eng. Osama Shaarawy

Revision midterm 2

Chapter 5: Some discrete probability distributions:

$\boxed{1}$ Discrete Uniform distribution:

discrete r.v. X assumes values: $x_1, x_2, \cdots, x_K$
with equal probabilities

$$f(x) = \begin{cases} \frac{1}{K} & \text{if} \quad x = x_1, x_2, \cdots, x_K \\ 0 & \text{otherwise} \end{cases}$$

Mean $\mu_X = E(X) = \dfrac{\sum_{i=1}^{K} x_i}{K}$

Variance $\sigma_X^2 = Var(X) = \dfrac{\sum_{i=1}^{K} (X_i - \mu_X)^2}{K}$

$\boxed{2}$ Binomial distribution:

Bernoulli process:

- n independent trials
- each trial has two possibilities success(p) (interest of study), failure(q) $= 1-p$

X no. of successes in n trials

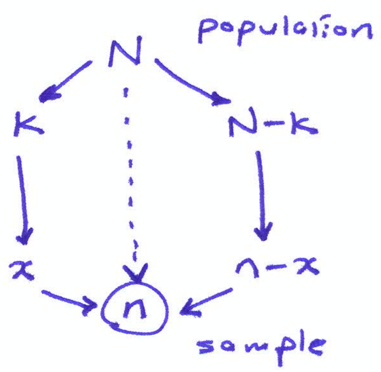
$$f(x) = \binom{n}{x} p^x q^{n-x} \quad ; \quad x = 0, 1, 2, \cdots, n$$

$P(X=x)$

Mean $\mu_X = np$

Variance $\sigma_X^2 = npq$

$\boxed{3}$ Hypergeometric distribution:



X no. of successes in n-size sample

→ selection with replacement

$X \sim$ binomial (n, p), $p = \frac{k}{N}$

→ selection without replacement

$$f(x) = \frac{\binom{K}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad ; \quad x = 0, 1, 2, \cdots, n$$

$X \sim$ hypergeometric $(N; K; n)$

Mean $\mu_X = n \cdot \frac{K}{N}$ Variance $\sigma_X^2 = n \cdot \frac{K}{N}\left(1-\frac{K}{N}\right)\frac{N-n}{N-1}$

note: hypergeometric is approximated to binomial distribution if: $n \ll N$, $n \ll K$

[4] Poisson distribution:

X ... no. of outcomes occuring in an interval t

$X \sim$ Poisson (μ) , $\mu = \lambda t$

λ: average no. of outcomes per unit.

$f(x) = e^{-\mu} \cdot \frac{\mu^x}{x!}$, $x = 0, 1, 2, \ldots$

Mean: $\mu_X = \mu = \lambda t$, Variance $\sigma_X^2 = \lambda t$

note: Binomial distribution is approximated to Poisson distribution if: $n \to \infty$, $p \to 0$

& $\mu = np$

Chapter 6:

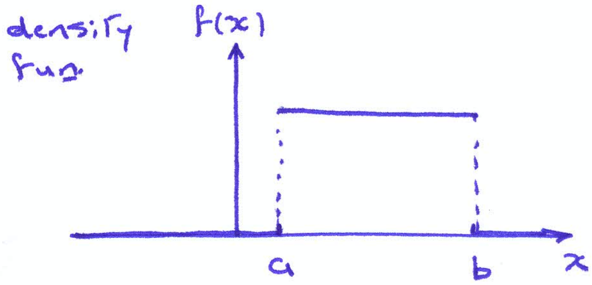
Some continuous probability distributions:

[1] Continuous p Uniform distribution:

$X \sim U(a,b)$

remember $P(X=x)=0$

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases}$$



Mean: $\mu_x = \frac{a+b}{2}$

Variance: $\sigma_x^2 = \frac{(b-a)^2}{12}$

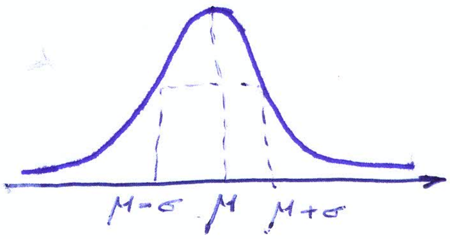
[2] Normal distribution:

$X \sim N(\mu, \sigma)$ mean μ, standard deviation σ

$\Downarrow$

standard normal distribution

$Z \sim N(0,1)$, $\boxed{Z = \frac{X-\mu}{\sigma}}$



$P(X \leq \square) = P(Z \leq \frac{\square - \mu}{\sigma})$

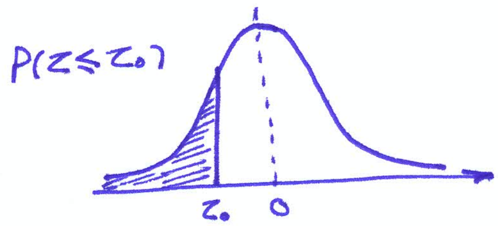
$P(X > \square) = 1 - P(X \leq \square)$

$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$

note that: percentage % = probability * 100

Total area under the curve = 1

$P(Z \leq 0) = P(Z > 0) = 0.5$



[3] Exponential distribution:

$$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \quad ; \quad x > 0$$

$$X \sim Exp(\beta)$$

mean $\mu_X = \beta$

variance $\sigma_X^2 = \beta^2$

Chapter 8: Fundamental Sampling distribution:

Random sampling:

Population $\equiv$ probability distribution of X

$\Downarrow$

random sample of size n

(independent & random observations $x_1, x_2, \dots, x_n$

parameter $\longrightarrow$ fun. of population

statistic $\longrightarrow$ fun. of sample

* sample mean $\bar{X} = \frac{\sum X_i}{n}$ (central tendency)
* sample variance $S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$ (variability)

$\longrightarrow$ Sampling distribution of $\bar{X}$

population $\quad X \sim N(\mu, \sigma)$

$\Downarrow$

sample $X_1, X_2, \dots, X_n$ $\quad \bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ $\quad \mu_{\bar{X}} = \mu$

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

$\downarrow$

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

for a large sample ($n \geq 30$)

$\bar{X}$ assumes a normal distribution.

$\longrightarrow$ sampling distribution of $\bar{X}_1 - \bar{X}_2$

1st population $\quad$ 2nd population

$X_1 \sim N(\mu_1, \sigma_1) \quad X_2 \sim N(\mu_2, \sigma_2)$

$\Downarrow n_1 \quad n_2 \Downarrow$

$\bar{X}_1 \sim N(\mu, \frac{\sigma_1}{\sqrt{n_1}}) \quad \bar{X}_2 \sim N(\mu_2, \frac{\sigma_2}{\sqrt{n_2}})$

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

$$Z \sim N(0,1) \qquad Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

note : $\mu_{\bar{X}_1 - \bar{X}_2} = E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$

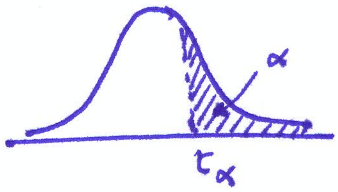
$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

→ t - distribution:

when σ^2 is unknown

A statistic $T = \dfrac{\bar{X} - \mu}{s/\sqrt{n}}$ $\qquad T \sim t(\nu)$

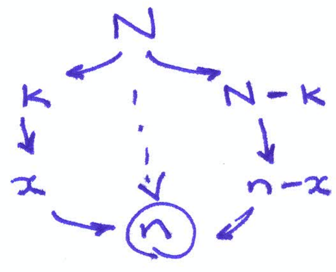
$\nu = n - 1$ degrees of freedom



$P(t > t_\alpha) = \alpha$

→ Sampling distribution of the sample proportion:

Population proportion $p = \frac{K}{N}$



Sample proportion $\hat{p} = \frac{x}{n}$

note: $X \sim$ binomial

$$\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$$ for large n

$$E(\hat{p}) = p \qquad Var(\hat{p}) = \frac{pq}{n}$$

sample same as population

→ Sampling distribution of $\hat{P}_1 - \hat{P}_2$

(difference between two proportions)

$$\hat{P}_1 = \frac{x_1}{n_1}, \quad \hat{P}_2 = \frac{x_2}{n_2}$$

$$\hat{P}_1 - \hat{P}_2 \sim N\left(P_1 - P_2, \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}\right)$$

Assume approximate normal distribution for large n_1, n_2

$$E(\hat{P}_1 - \hat{P}_2) = P_1 - P_2$$

$$Var(\hat{P}_1 - \hat{P}_2) = \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}$$