conditional probability:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Permutations & Combinations:

Permutations: ${}_nP_r = \frac{n!}{(n-r)!}$ "arranged"

Combinations: ${}_nC_r = \frac{n!}{r!(n-r)!}$: order doesn't matter

- If operation can be performed in n_1 ways & for each of these ways, consequent operations can be done in n_2, n_3, n_k ways, then the sequence of k operations can be performed in $n_1 n_2 \cdots n_k$ ways.

- The number of distinct permutations of n things of which n_1 are of one kind, n_2 of a second kind $\cdots n_k$ of a k_{th} kind: $\frac{n!}{n_1!\,n_2!\,n_k!}$

L7: Cont distributions

1. $f(x) \geq 0$, 2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

3. $P(a < X < b) = \int_a^b f(x)\,dx$

$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt \quad -\infty < x < \infty$

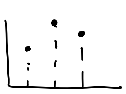
e.g. finding $F(x)$

$f(x) = \frac{x^2}{3} \quad -1 < x < 2$

$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{-1}^{x} \frac{t^2}{3} = \left.\frac{t^3}{9}\right]_{-1}^{x}$

$$F(x) = \begin{cases} 0 & x < -1 \\ \frac{x^3+1}{9}, & -1 \leq x < 2 \\ 1, & x \geq 2 \end{cases} \quad = \frac{x^3+1}{9}$$

Discrete RV. Prob mass function:



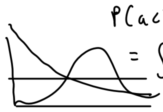
$P(X = x)$

Cumulative:

$P(X \leq x)$

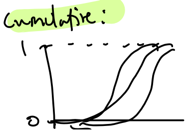
Cont RV:

Prob density func:



$P(a < x < b) = \int_b^a f(t)\,dt$

Cumulative:



$F(x) = \int_{-\infty}^{x} f(t)\,dt$

L8: Joint distributions:

1. $f(x,y) \geq 0 \ \forall (x,y)$, 2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

3. $P((X,Y) \in A) = \iint_A f(x,y)\,dx\,dy$

Marginal distributions:

$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$ & $h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$

conditional probability distribution:

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f(x,y)}{g(x)}$$

independence:

if $f(x|y)$ does not depend on y, then $f(x|y) = g(x)$ & $f(x,y) = g(x)\,h(x)$

- needs to be for all values of (x,y)

- Random variables $X_1 \cdots X_n$ are mutually statistically independent iff

L9: expectation & variance

$E(X)$ or $\mu_x = \int_{-\infty}^{\infty} x f(x)\,dx$ or $\int_{-\infty}^{\infty} g(x) f(x)\,dx$

Expectation of joint distribution:

$$E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\,f(x,y)\,dx\,dy$$

variance (σ^2):

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

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

$= E((g(x) - \mu_{g(x)})^2)$

$\sigma_{xy} = \mathrm{cov}(X,Y)$

$= E((X-\mu_x)(Y-\mu_y))$

$= E(XY) - E(X)E(Y)$

Correlation coefficient of X & Y:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \quad -1 \leq \rho_{xy} \leq 1$$

Properties of expectation:

$E(b) = b$, $E(aX) = aE(X)$, $E(aX+b) = aE(X)+b$

$E(X+Y) = E(X) + E(Y)$

$E(XY) = E(X)E(Y)$ only if X & Y are indep.

Properties of Variance:

$\mathrm{Var}(a) = 0$, $\mathrm{var}(X+b) = \mathrm{var}(X)$

$\mathrm{var}(aX) = a^2\,\mathrm{var}(X)$

$\mathrm{var}(aX \pm bY) = a^2\,\mathrm{var}(X) + b^2\,\mathrm{var}(Y) \pm 2ab\,\mathrm{cov}(X,Y)$

Chebyshev's Theorem:

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

e.g. $\mu = 8$ $\sigma^2 = 9$; find Chebyshev's bounds:

$P(|X-8| \geq 6) = 1 - P(|X-8| < 6)$

$= 1 - P(-6 < x-8 < 6)$

$= 1 - P(2 < x < 14)$

$|X-8| < 6$

$\rightarrow x - 8 < 6$

$-x - 8 < 6$

$\rightarrow -6 < x-8 < 6$

$8 - k3 = 2 \rightarrow k = 2$

$\rightarrow \ = P(8 - 2(3) < x < 8 + 2(3)) \geq 1 - \frac{1}{2^2}$

so $\therefore$ $1 - P(2 < x < 14) \leq \frac{1}{4}$

Binomial distribution:

$X \sim B(n,p) \rightarrow b(x; n, p) = P(X = x)$

$= \binom{n}{x} p^x q^{n-x}$

$P(X \leq r) = P(X=0) + P(X=1) + \cdots + P(X=r)$

$P(3 \leq X \leq 8) = P(X \leq 8) - P(X \leq 2)$

$= B(8; n, p) - B(2; n, p)$

$E(X) = np$, $\mathrm{Var}(X) = npq$

Multinomial distribution:

k pos. outcomes of events $E_1 \cdots E_k$ with prob. $p_1 \cdots p_k$,

$P(X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k) =$

$\binom{n}{x_1 \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$, where

$\binom{n}{x_1, x_2 \cdots x_k} = \frac{n!}{x_1!\,x_2! \cdots x_k!}$, $\sum x_i = n$, $\sum p_i = 1$

Hypergeometric distribution:

- not Bernoulli, - no replacement

Let X = RV, representing # of successes in sample of n, n drawn from N, K successes,

$$P(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

if $\frac{n}{N} < 0.05$, approx using binomial dist. $\Rightarrow X \sim B(n = n, p = \frac{K}{N})$

$E(X) = \frac{nK}{N}$

$\sigma^2 = \left[\frac{N-n}{N-1}\right](n)\left(\frac{K}{N}\right)\left(\frac{N-K}{N}\right)$

Multivariate hypergeometric:

- N partitioned into $A_1 \cdots A_m$ with $a_1 \cdots a_m$ elements.

- sample size is n

- how many of each element, $x_1, x_2 \cdots x_m$

$P = (X_1 = x_1, X_2 = x_2, \cdots, X_m = x_m) =$

$$\frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_m}{x_m}}{\binom{N}{n}}, \quad \sum x_i = n, \quad \sum a_i = N$$

negative binomial:

- "How many trials for k successes?"

- let X be the RV rep. # of trials to get k successes.

$P(X = x) = \binom{x-1}{k-1} p^k q^{x-k}$

- last trial must be a success.

Geometric dist.

- special case of negative bin. prob of success = p.

- X is a RV representing # of trials until first success. $P(X = x) = (1-p)^{x-1} p$

- $F(x) = 1 - (1-p)^x$

$E(X) = \frac{1}{p}$ $\quad \sigma^2 = \frac{1-p}{p^2}$

Poisson dist:

convert to time

- n trials, x successes, p prob successes

t time, x successes, λ arrival rate of success

X is RV of # of successes in t time, where success arrive with rate λ.

$$P(X=x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

$$\mu = \sigma^2 = \lambda t$$

Properties:

independence: # of arrivals in a time interval independent of # of outcomes in any other disjoint time interval.

Time space homogenity: prob of k arrivals in t time is the same for all intervals of length t.

small interval probabilities: prob that 2 or more arrivals occur in a small time interval is negligible.

Question 3 (8 marks)
Metalmec is a company that manufactures aluminum casts. The time in hours that a die cast takes to solidify and cool can be modelled as a random variable X with the following probability density function:

$$f(x) = \begin{cases} kx, & 0 \le x < 2 \\ 1 - \frac{x}{4}, & 2 \le x < 4 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the value of k so that $f(x)$ is correctly defined. (5 marks)

Solution:

$$\int_{-\infty}^{\infty} f(t)\,dt = 1$$

$$\int_0^2 kt\,dt + \int_2^4 \left(1 - \frac{t}{4}\right) dt = 1$$

$$\left.\frac{kt^2}{2}\right|_0^2 + \left.\left(t - \frac{t^2}{2(4)}\right)\right|_2^4 = 1$$

$$2k + \frac{1}{2} = 1$$

$$k = \frac{1}{4}$$

(b) Give an expression for the cumulative distribution function $F(x)$ (3 marks)

Solution:
For $x < 2$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x \frac{1}{4} t\,dt = \left.\frac{t^2}{8}\right|_0^x = \frac{x^2}{8}$$

For $x \le x < 4$

$$F(x) = \left.\frac{t^2}{8}\right|_0^2 + \int_2^x \left(1 - \frac{t}{4}\right) dt = \frac{1}{2} + \left.\left(t - \frac{t^2}{8}\right)\right|_2^x = x - \frac{x^2}{8} - 1$$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{8} & 0 \le x < 2 \\ x - \frac{x^2}{8} - 1 & 2 \le x < 4 \\ 1 & x \ge 4 \end{cases}$$

Phrase	Means	Calculation
… greater than 5 …	$X > 5$	$1 - P(X \leqslant 5)$
… no more than 3 …	$X \leqslant 3$	$P(X \leqslant 3)$
… at least 7 …	$X \geqslant 7$	$1 - P(X \leqslant 6)$
… fewer than 10 …	$X < 10$	$P(X \leqslant 9)$
… at most 8 …	$X \leqslant 8$	$P(X \leqslant 8)$

x	1	2	3	4	5
$F(x)$	0.2	0.32	0.67	0.9	1

$$P(X=3) = F(3) - F(2) = 0.67 - 0.32 = 0.35$$

$$P(X > 2) = 1 - F(2) = 1 - 0.32 = 0.68$$

$$P(2 \le x \le 4) = F(4) - F(1) = 0.9 - 0.2 = 0.7$$

Cumulative example

- A car agency sells 50% of its inventory with side airbags. Find the formula for the probability distribution of the number of cars with side airbags among the next 4 cars sold.

sample space $= 2 \times 2 \times 2 \times 2 = \underline{16}$

for x cars with airbags,

$$f(x) = \frac{\binom{4}{x}}{16}, \quad x = 0, 1, 2, 3, 4$$

$$F(x) = P(X \le x) = \sum_{t \le x} f(t) \quad -\infty < x < \infty$$

- Find cumulative distribution function of X in side airbag problem.

$$f(x) = \frac{\binom{4}{x}}{16}, \quad x = 0 \ldots 4, \quad f(0) = \frac{1}{16}, \quad f(1) = \frac{1}{4}$$

$$f(2) = \frac{3}{8}, \quad f(3) = \frac{1}{4}, \quad f(4) = \frac{1}{16}$$

$F(0) = f(0) = \frac{1}{16}$

$F(1) = f(0) + f(1) = \frac{5}{16}$

$F(2) = f(0) + f(1) + f(2) = \frac{11}{16}$

$F(3) = f(0) + \cdots + f(3) = \frac{15}{16}$

$F(4) = 1$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{16} & 0 \le x < 1 \\ 5/16 & 1 \le x < 2 \\ 11/16 & 2 \le x < 3 \\ 15/16 & 3 \le x < 4 \\ 1 & x \ge 4 \end{cases}$$

$$f(2) = F(2) - F(1) = \frac{3}{8}$$

$$F(2) = f(0) + f(1) + f(2) = F(1) + f(2)$$

Continuous uniform dist.

- $f(x) = \frac{1}{B - A}, \quad A \le x \le B$

$$\mu = \frac{A+B}{2}, \quad \sigma^2 = \frac{(B-A)^2}{12}$$

Normal dist.

- density of the random variable X w mean μ & var σ^2:

Normal Distribution The density of the normal random variable X, with mean μ and variance σ^2, is

$$n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty,$$

Properties: - area under curve = 1
- mode occurs at μ, - symmetric through μ
- point of inflection at $x = \mu \pm \sigma$

$P(a < x < b)$ = area under curve between (a, b)

$$P(a < x < b) = \int_a^b n(x; \mu, \sigma)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx$$

$$N(\mu, \sigma^2) \rightarrow N(0, 1) \text{ (standardizing)}$$

$$z = \frac{X - \mu}{\sigma} \leftarrow \text{default notation}$$

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right)$$

$x = \sigma z + \mu$ ← always look at area to left when evaluating x & z values.

If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$, then the limiting form of the distribution of

$$Z = \frac{X - np}{\sqrt{npq}},$$

as $n \to \infty$, is the standard normal distribution $n(z; 0, 1)$.

np & $nq \ge 5$
- n is large
- p close to $\frac{1}{2}$

$$P(X \le x) = \sum_{k=0}^{x} b(k; n, p) \approx P\left(Z \le \frac{x - np + 0.5}{\sqrt{npq}}\right)$$

$$P(a \le X \le b) \approx \left(\frac{a - np - 0.5}{\sqrt{npq}} \le z \le \frac{b - np + 0.5}{\sqrt{npq}}\right)$$

- The ± 0.5 is a continuity correction

Exp & Gamma dist.

Exponential distribution:
"probability an event will occur by time t."
$= P(T \le t)$

$$f(t) = \lambda e^{-\lambda t} \quad \text{or} \quad \frac{1}{\beta} e^{-\frac{1}{\beta} t}$$

λ is β, called "rate"

CDF: $F(t) = P(T \le t) = 1 - e^{-\lambda t}$ ← prob arrival occurs $[0, t]$

$P(T > t) = e^{-\lambda t}$ ← same as Poisson when $x = 0$.
"Time to first arrival $T \ge t$"

$$\mu = \frac{1}{\lambda} \qquad \sigma^2 = \frac{1}{\lambda^2}$$

$$E(X \mid a) = a + \frac{1}{\lambda}$$

Gamma distribution:

"prob of waiting x time until α events occur"

Gamma function (not a dist.):

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx, \quad \alpha > 0$$

Properties of Gamma function

a) $\Gamma(n) = (n-1)(n-2)(n-3)\cdots(1)$ for $n>0$, $n \in \mathbb{Z}$

b) $\Gamma(n) = (n-1)!$ $\quad n>0, n\in\mathbb{Z}$

c) $\Gamma(1) = 1$

d) $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

shape & rate version: PDF

$$f(x) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x>0$$

'λ' is rate, $\mu = \frac{\alpha}{\lambda}$, $\sigma^2 = \frac{\alpha}{\lambda^2}$

$$F(x;\alpha) = \int_0^\infty \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y}\, dy$$

CDF $\quad y = \lambda x$, $x = \frac{y}{\lambda}$, $dx = \frac{1}{\lambda} y$

Chi-Squared distribution

- take gamma dist., set $\alpha = \frac{v}{2}$, $\lambda = \frac{1}{2}$, we get the chi-squared dist.

χ^2 PDF: $f(x) = \dfrac{x^{\frac{v}{2}-1} e^{-x/2}}{2^{v/2}\,\Gamma(\frac{v}{2})}, \quad x>0$

$\mu = v$, $\sigma^2 = 2v$

'v' is called "degrees of freedom"

Functions of random variables:

Transforming discrete RV:

$Y = u(X)$ $\quad$ 1:1 mapping between X & Y

$\therefore y = u(x) \rightarrow x = w(y)$

$\therefore$ PDF of Y: $g(y) = f[w(y)]$

Transforming 2D discrete RV:

Suppose we have two independent RV's X_1, X_2. Find PDF of $Y_1 = X_1 + X_2$

→ Introduce second mapping: $Y_2 = X_2$

$\therefore Y_1 = X_1 + X_2 \rightarrow X_1 = Y_1 - X_2 = Y_1 - Y_2$

$Y_2 = X_2 \rightarrow X_2 = Y_2$

→ replace (x_1, x_2) with (y_1, y_2) & get marginal of y_1

$$g(y_1,y_2) = f[w_1(y_1,y_2), w_2(y_1,y_2)]$$

transformation of continuous RV:

- get $w(y)$ which is x in terms of y.
- Find Jacobian which is $w(y)'$.
- sub $w(y)$ into PDF & multiply by J.

$$\rightarrow g(y) = f[w(y)]\,|J|$$

transformation of continuous RV with multi-variables:

$$g(y_1,y_2) = f[w_1(y_1,y_2), w_2(y_1,y_2)]\,|J|$$

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

Sampling Distributions:

- variance of sample of size n:

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

- The prob dist. of a statistic is a sampling distn.
- Suppose we have population of mean $= \mu$, variance $= \sigma^2$.

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

variance of sample mean is population variance divided by n.

Central Limit theorem:

If $n \geq 30$, CLT is good aproximation

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ as $n \rightarrow \infty$

limiting distribution:

$Z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ as $n \rightarrow \infty$ is standard normal

Sampling distribution of difference between two means:

Pop 1: mean μ_1, variance σ_1^2, sample mean $\bar{X}_1$

Pop 2: mean μ_2, variance σ_2^2, sample mean $\bar{X}_2$

By CLT, $\bar{X}_1$ & $\bar{X}_2$ are normal with mean μ_1, μ_2 & variance $\frac{\sigma_1^2}{n_1}, \frac{\sigma_2^2}{n_2}$

for distribution of $\bar{X}_1 - \bar{X}_2$, we need mean & variance of $\bar{X}_1 - \bar{X}_2$.

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

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

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

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

Sampling distributions of s^2:

- For sample variance, we use chi-squared dist.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \sum \frac{(x_i - \bar{x})^2}{\sigma^2}$$

given in question

- $v = n-1$ degrees of freedom
- s^2 = variance
- evaluate critical vals for chi-squared distribution at 0.975 & 0.025, then determine if χ^2 falls in this range to see if claim in question is correct.

t-distribution:

We don't know σ in reality, so we use s as a proxy and use student "t-distribution"

$T = \dfrac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$, $\bar{X} = \frac{1}{n}\sum X_i$

$s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$

- $n-1$ degrees of freedom