STAT 230 : Probability **Spring 2024**
Formula / Theorem Sheet **by Strafe <3**

1 Chapter 1

- Classical Definition $\dfrac{\text{Number of ways it can occur}}{\text{Total number of outcomes}}$
- Relative Frequency (Portion or fraction of times it can occur in a very long series of repetition)
- Surjective probability (How sure the person making the statement is)

2 Chapter 2

A sample space is discrete if it is finite Simple event - an event containing only one point Compound even - an event made up of two or more simple events

2.1 Odds

Odds in favour of an event A

$$\text{in favour} = \frac{P(A)}{1 - P(A)}$$

Odds against an event

$$\text{against} = \frac{1 - P(A)}{P(A)}$$

2.2 Chapter 3

```
if (selection with replacement) {
    n^k
} else {
    if (order matters) {
        permutation (arrangements or lists)
    } else {
        combination (subsets)
    }
}
```

2.3 Permutation

$$n^{(k)}$$

2.4 Combination

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

2.5 Chapter 4

2.6 De Morgan's Law

$$(A \cup B)^C = A^C \cap B^C$$
$$(A \cap B)^C = A^C \cup B^C$$

2.7 Rules

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

2.8 Independence

A and B are independent iff

$$P(A \cap B) = P(A)P(B)$$

2.9 Mutually Exclusive

$$P(A \cap B) = 0$$

2.10 Conditional

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

2.11 Product Rule

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

2.12 Bayes Theorem

$$P(B) = P(B \cap A_1) + \cdots + P(B \cap A_k)$$
$$P(B) = P(B)P(A_1|B) + \ldots P(B)P(A_k|B)$$

3 Chapter 5

3.1 Discrete Uniform Distribution

$$f(x) = \frac{1}{b-a+1}, x = a, a+1, \ldots, b$$

3.2 Hypergeometric Distribution

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

$\min(r, n) \geq x \geq \max(0, n-(N-r))$
N objects in total where r are successes, and we randomly select n without replacement
X = number of successes observed among n trials

$$X \sim Hypergeo(N, r, n)$$

3.3 Binomial Distribution

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \ldots, n$$

Requirements:

- Two Trials
- Independent Trials (With replacement)
- Multiple Trials
- Same Probability

X = number of successes observed among n trials

$$X \sim Binom(n, p)$$

3.4 Negative Binomial Distribution

$$f(x) = \binom{x+k-1}{x} p^k (1-p)^x, x = 0, 1, 2, \ldots$$

Two outcomes, independent trials, same probability of success. Do experiments until we obtain k successes. X = the number of **failures** obtained before k^{th} success.

$$X \sim NB(k, p)$$

3.5 Geometric Distribution

$$f(x) = p(1-p)^x, x = 0, 1, 2, \ldots$$

Two outcomes, independent trials, same probability of success. X = the number of **failures** obtained before 1st success.

$$X \sim Geo(p)$$

3.6 Poisson Distribution

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

X = the number of events of some type. The events occur according to $\mu = np$.

3.7 Poission Distribution from Poison Process

Order notation:

$$g(\Delta t) = o(\Delta t), \Delta t \to 0$$

This means g approaches 0 quicker than Δt as Δt approaches 0.

$$\frac{g(\Delta t)}{\Delta t} \to 0$$

Setup: assume a certain type of event occurs at random points in time and satisfies the following

- **Independence** The number of occurrences in non-overlapping intervals are independent
- **Individuality** P(2 or more events in $(t, t+\Delta t)) = o(\Delta t)$ as $t \to 0$. (Events do not occur in clusters, but individually)
- **Homogeneity / Uniformity** Occurs at a uniform rate λ over time so

$$P(\text{one event in } (t, t+\Delta t)) = \lambda \Delta t + o(\Delta t)$$

$$f(x) = P(X = x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, x = 0, 1, 2, \ldots$$

X =number of events in a specified length

$$X \sim Poi(\lambda), \mu = \lambda t$$

Also applies when events occur randomly in space (replace time with volume or area)

4 Chapter 7

4.1 Frequency Distribution

- Mean $>$ Median (right-skewed)
- Mean $=$ Median (Symmetric)
- Mean $<$ Median (Left-Skewed)

4.2 Arithmetic Mean

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

4.3 Expected Value

Expected value (mean or expectation) of a **discrete** random variable X with probability function $f(x)$ is:

$$\mu = E(X) = \sum_{allx} x f(x)$$

4.4 Properties of Expectation

$$E[ag(X) + b] = aE[g(X)] + b$$

$$E[Ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)]$$

$$E(a) = a, E[g(X)] = g(E[X]) \text{ only if g(X) is a linear function}$$

4.5 Expected Values of Distributions

- Binomial $E(X) = np$
- Poisson $E(X) = \mu = \lambda t$
- Discrete $E(X) = \dfrac{a+b}{2}$
- Hypergeometric $E(X) = \dfrac{nr}{N}$
- Negative Binomial $E(X) = \dfrac{k(1-p)}{p}$
- Geometric $E(X) = \dfrac{1-p}{p}$

4.6 Interquartile Range (IQR)

$$IQR = Q_3 - Q1, \tilde{x} = Q2$$

4.7 Variability

$$E(X-\mu) = E(X) - \mu = 0$$
$$\sigma^2 = Var(X) = E[(X-\mu)^2]$$
$$\sigma = \sqrt{E[(X-\mu)^2]}$$
$$Var(X) = E(X^2) - \mu^2 = E[X(X-1)] + \mu - \mu^2$$

4.8 Variance of Distributions

- Binomial $Var(X) = np(1-p)$
- Poisson $Var(X) = \mu^2$
- Discrete Uniform $Var(X) = \dfrac{(b-a+1)^2-1}{12}$
- Hypergeometric $Var(X) = \dfrac{nr}{N}(1-\dfrac{r}{N})\dfrac{N-n}{N-1}$
- Negative Binomial $Var(X) = \dfrac{k(1-p)}{p^2}$
- Geometric $Var(X) = \dfrac{1-p}{p^2}$

4.9 Properties of Mean and Variance

Let $Y = aX + b$, $E(a) = a, Var(a) = 0$

1. $\mu_Y = E(Y) = aE(X) + b$
2. $\sigma_Y^2 = Var(Y) = a^2Var(X)$

5 Chapter 8

5.1 Cumulative Distribution Function (c.d.f)

- $F(x) = P(X \leq x)$
- $F(x)$ is defined for all $x \in \mathbb{R}$
- $\lim_{x \to \infty} F(x) = 0, \lim_{x \to \infty} F(x) = 1$
- $F(x)$ is a non decreasing function
- $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$
- $P(X = x) = 0, P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = F(b) - F(a)$

5.2 Probability Density Function (p.d.f)

Suppose an interval $[x, x + \Delta x]$

$$P(x \leq X \leq x + \Delta x) = F(x + \Delta x) - F(x)$$

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x} = \frac{dF(x)}{dx} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

Continuous r.vs.

$$f(x) = \frac{d}{dx} F(x), F(x) = P(X \leq x) = \int_{-\infty}^{x} f_x(x)dx$$

Discrete r.vs.

$$f(x) = P(X = x) = F(x) - F(x-1),\ F(x) = \sum_{w=-\infty}^{x} f(w)$$

Properties:

- $P(a \leq X \leq b) = F(b) - f(a) = \int_a^b f(x)dx$
- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = \int_{all x} f(x)dx = 1$
- $F(x) = \int_{-\infty}^{x} f(u)du$
- $P(x - \frac{\Delta x}{2} \leq X \leq x + \frac{\Delta x}{2}) = F(x + \frac{\Delta x}{2}) - F(x - \frac{\Delta x}{2}) \simeq f(x)\Delta x$

5.3 Expectation, Mean, Variance of Continuous Distribution

$$E(g(X)) = \int_{allx} g(x)f(x)dx$$

$$E(x) = \int_{allx} xf(x)dx$$

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

5.4 Continuous Uniform Distribution

X is a r.v. taking on values in the interval $[a, b]$ with all subintervals of a fixed length being equally likely

$$X \sim Unif[a, b]$$

The probably density function (p. d. f) must be a constant $f(x) = k$ so

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & otherwise \end{cases}$$

The Cumulative Distribution function

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

$$E(X) = \frac{b+a}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

5.5 Exponential Distribution

In a Poisson Process for events in time, then the length of time we wait until the first occurence.
λ = average number of occurrences per unit of time
θ = average waiting time for an occurrence

$$F_x(x) = 1 - e^{-x/\theta}, x > 0$$

$$f_x(x) = \frac{1}{\theta}e^{-x/\theta}, x > 0$$

$$E(x) = \theta = \frac{1}{\lambda}, Var(X) = \theta^2 = \frac{1}{\lambda^2}$$

5.6 Gamma Function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x} = (\alpha - 1)!$$

5.7 Memoryless Property

For exponential Distributions,

$$P(X > c + b | X > b) = P(X > c)$$

5.8 Normal Distribution

A r.v. defined on $(-\infty, \infty)$ has a normal distribution if it has form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

$-\infty < \mu < \infty$, $\sigma > 0$
μ shifts distribution along x axis, σ^2 stretches or pulls the distribution

5.9 Finding Probabilities on N(0, 1) tables

Let $X \sim N(\mu, \sigma^2)$

$$Z = \frac{X - \mu}{\sigma}$$

Then $Z \sim (0, 1)$ (use table)

5.10 Gaussian Distribution

$$X \sim G(\mu, \sigma)$$

6 Chapter 9

6.1 Joint Probability Function

$$f(x, y) = P(X = x, Y = y)$$
$$f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

if there are n r.v.s $X_1, \ldots, X_n$

- $f(x, y) \geq 0$ for all (x, y)
- $\sum_{all(x,y)} f(x, y) = 1$

6.2 Marginal Distribution

$$f_1(x) = f_X(x) = \sum_{ally} f(x, y)$$
$$f_2(y) = f_Y(y) = \sum_{allx} f(x, y)$$
$$f_{1,3}(x_1, x_3) = \sum_{allx_2} f(x_1, x_2, x_3)$$

6.3 Independent R.V.

$$f(x, y) = f_1(x)f_2(y) \iff \text{It is independent}$$

$f(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n)$ also applies

6.4 Conditional Probability

$$f(x|y) = \frac{f(x, y)}{f_2(y)} = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$P(Y = y) > 0$. If it is independent, $= f_1(x)$

6.5 Functions

If $x + y = t, y = t - x$

$$f_T(t) = P(T = t) = \sum_{allx} f(x, t - x) = \sum_{allx} P(X = x, Y = t - x)$$

In general, $U = g(X_1, \ldots X_n)$ of two R.V.s X, Y

$$F_U(u) = P(U = u) = \sum_{all(x_1,\ldots x_n), g(x_1,\ldots x_n) = u} f(x_1, \ldots x_n)$$

If $X \sim Bin(n, p)$ and $Y \sim Bin(m, p)$ **independently** then

$$T = X + Y \sim Bin(n + m, p)$$

6.6 Multinomial Distribution

The experiment is repeated independently n times with k distinct outcomes. Let the probability of these k types be $p_1, p_2, \dots p_k$ each time. Let X_i be the. number of times be i-th type occurs

- $p_1 + \dots + p_k = 1$
- $X_1 + \dots + X_k = n$
- $\sum f(x_1, \dots x_k) = 1$

$$f(x_1, \dots, x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x}2 \dots p_k^{x_k}$$

If we are only interested in X_2,
$P_2 = 1 - P_1 - \dots - P_k$ and $X_2 \sim Bin(n; p_2)$
If $T = X_1 + X_2, T \sim Bin(n; p_1 + p_2)$
$X_1 | T = t \sim Bin\left(t; \frac{P_1}{P_1 + P_2}\right)$

6.7 Covariance and Correlation

Recall $E[g(x)] = \Sigma_{allx} g(x) f(x)$

$$E[g(X, Y)] = \Sigma_{all(x,y)} g(x, y) f(x, y)$$

$$E[ag_1(X, Y) + bg_2(X, Y)] = aE[g_1(X, Y)] + bE[g_2(X, Y)]$$

Covariance

$$Cov(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E(X)E(Y)$$

$\mu_X = E(X)$

6.8 Independence and Covariance

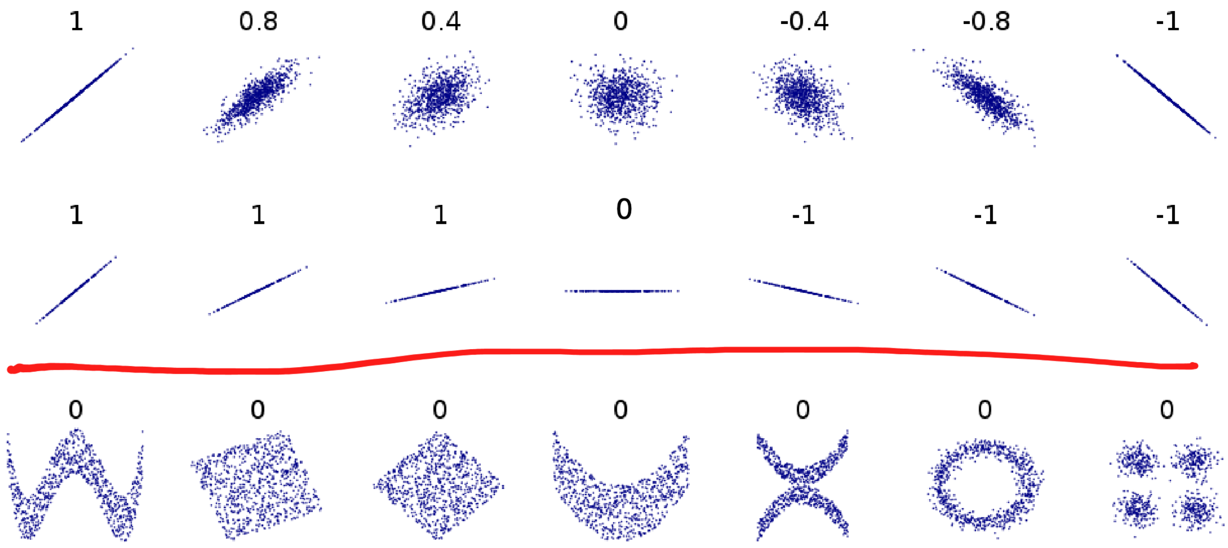
If X, Y are independent, $Cov(X, Y) = 0$ If X, Y are independent R.V.s

$$E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$$

6.9 Correlation Coefficient

$$\rho = Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Lies within [-1, 1]. It will have the same sign as Cov(X, Y). As $\rho \to \pm 1$ the relationship between X, Y becomes closer to linear.



Covariance interpret the sign, **Correlation** interpret the magnitude AND sign

6.10 Expectation

- $E(aX + bY) = aE(X) + bE(Y)$
- If $E(X_i) = \mu_i$), $E(\Sigma a_i X_i) = \Sigma a_i \mu_i$
- Let $X_1, \ldots, X_n$ have mean μ,

$$E(\frac{\Sigma_{i=1}^n x_i}{n}) = \mu$$

1. $Cov(X, X) = E[(x - \mu_X)(x - \mu_X)] = E[(x - \mu_X)^2] = Var(x)$
2. $Cov(aX + bY, cU + dV) = acCov(X, U) + adCov(X, V) + bcCov(Y, U) + bdCov(Y, V)$

6.11 Results for Variance

1. Variance of a linear combination for r.v.s X, Y and constants a,b

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

2. Variance of a sum of independent r.v.s, since $Cov(X, Y) = 0$

$$Var(X + Y) = Var(X - Y) = \sigma_X^2 + \sigma_Y^2$$

3. Variance of a general linear combination. Let a_i be constants and $Var(X_i) = \sigma_i^2$

$$Var(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 \sigma_i^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n a_i a_j Cov(X_i, X_j)$$

4. Variance of a linear combination of independent random variables

   (a) If $X_1, X_2, \ldots, X_n$ are independent random variables, then $Cov(X_i, X_j) = 0$ so that

   $$Var(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i)$$

   (b) If $X_1, X_2, \ldots, X_n$ are independent r.v.s and all have the same variance σ^2 then

   $$Var(\frac{\sum_{i=1}^{n} X_i}{n}) = \frac{\sigma^2}{n}$$

6.12 Linear Combination of Independent Normal R.V.

Let $X \sim N(\mu, \sigma^2), Y = aX + B$,

$$y \sim N(a\mu + b, a^2\sigma^2)$$

Let $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ be independent R.V.s

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

Let $X_1, \ldots, X_n$ be independent $N(\mu, \sigma^2)$ random variable.

$$\text{Total} = \sum_{i=1}^{n} X_i \sim N(n\mu; n\sigma^2)$$

$$\text{Sample Mean} = \overline{x} = \frac{\sum_{i=1}^{n} X_i}{n} \sim N(\mu; \frac{\mu^2}{n})$$

6.13 Indicator Variables

Binary (0, 1) that indicates if an event has taken place. e.g. $X \sim Bin(n, p)$

$$X_i = \begin{cases} 0 & \text{if } i^{th} \text{ trial was a failure (probability 1 - p)} \\ 1 & \text{if } i^{th} \text{ trial was a success (probability p)} \end{cases}$$

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

$$E(X) = np$$

$$Var(x) = np(1-p)$$

7 Chapter 10

7.1 Central Limit Theorem (C.L.T.)

If $X_1, X_2, \ldots, X_n$ are **independent** r.v.s all have the same distribution mean μ and variance σ^2 then as $n \to \infty$, the cumulative distribution function of the random variable

$$\frac{(\sum_{i=1}^{n} X_i) - n\mu}{\sigma\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

approaches the $N(0,1)$ cumulative distribution function. Similarly the c.d.f. of

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

approaches the $N(0,1)$ c.d.f.

- If X_i themselves have a normal distribution, then $S_n, \overline{X}$ have exactly normal distribution for all values of n
- If X_i do not have a normal distribution, then $S_n, \overline{X}$ have approximately normal distribution for large values of n.

7.2 Normal Approximation to Poisson Distribution

Let $X \sim Poisson(\mu = \lambda t)$

$$Z = \frac{X - \mu}{\sqrt{\mu}}$$

is approximately $N(0,1)$

7.3 Normal Approximation to Binomial Distribution

Let $X \sim Bin(n,p)$ Then for n large, the r.v. $X \sim N(\mu = np; \sigma^2 = np(1-p))$

$$W = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately $N(0,1)$

7.4 Moment Generating Function (m.g.f)

$$M_x(t) = E[e^{tX}] = \sum_x e^{tx} f(x)$$

The m.g.f. is assumed to be defined and finite for values of $t \in [-a, a]$ for $a > 0$

7.5 m.g.f. Theorem

Let r.v. X have m.g.f. $M(t)$

$$E[X^r] = M^{(r)}(0), r = 1, 2, \ldots$$