ACTL2131 Course Notes

Lecturer Email:

Tutor Email:

Mathematical Models

Sample Space & σ-algebra

- The set of all possible outcomes is called the sample space Ω
- A family of subsets of the sample space is called the σ-algebra if:
  1. It contains the Sample Space: $\Omega \in \mathcal{F}$.
  2. It is closed under the complement operation: if $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$;
  3. It is closed under the union operation: if $E_1, E_2, \ldots$ are in $\mathcal{F}$, then $\bigcup_{k=1}^{\infty} E_k \in \mathcal{F}$.
- If there are N elements in the sample space, a σ-algebra may consist of up to 2^N elements.

Probability Measure

- A Probability measure is a function that satisfies the following:
  1. For all events E_i, $0 \leq \Pr(E_i) \leq 1$;
  2. $\Pr(\Omega) = 1$;
  3. For events $E_1, E_2, \ldots$ such that $E_i \cap E_j = \emptyset$ (for every $i \neq j$), we have:

$$\Pr\left(\bigcup_{k=1}^{\infty} E_k\right) = \Pr(E_1) + \Pr(E_2) + \ldots = \sum_{k=1}^{\infty} \Pr(E_k).$$

- A random experiment is described as a probability triple
- Properties of the Probability Measure:
  - **Union of two events**: the event that *A* and/or *B* occurs
  - **Intersection of two events**: the event that *A* and *B* occurs
  - **Complement of an event**: the event that *A* does not occur
  - **Events of disjoint**: If the events have no outcomes in common
- Useful Laws:
  - Commutative Laws

    $A \cup B = B \cup A$
  - Associative Laws

    $(A \cup B) \cup C = A \cup (B \cup C)$
  - Distributive Laws

    $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$
  - DeMorgans Law

    $(A \cap B)^c = A^c \cup B^c$ and $(A \cup B)^c = A^c \cap B^c$.

Inequality Rules

- Boole's Inequality:

$$\Pr\left(\bigcup_{k=1}^{n} E_k\right) \leq \sum_{k=1}^{n} \Pr(E_k)$$

- Bonferroni's Inequality:

$$\Pr(E_1 \cap \ldots \cap E_n) \geq 1 - \sum_{k=1}^{n} \Pr(E_k^c).$$

Conditional Probability

- The conditional probability of A given B is:

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

- The **multiplication rule** follows:

$$\Pr(A \cap B) = \Pr(A|B) \cdot \Pr(B)$$

- Properties:
  1. $\Pr(A|B) \geq 0$;
  2. $\Pr(A|A) = 1$;
  3. If $A_1, A_2, \ldots$ are mutually disjoint events, then $\Pr\left(\bigcup_{k=1}^{\infty} A_k |B\right) = \sum_{k=1}^{\infty} \Pr(A_k|B)$.

Law of Total Probability

- If $E_1, E_2, \ldots$ are mutually disjoint, countable and measurable, then for any event $A \in F$:

$$\Pr(A) = \sum_{k=1}^{\infty} \Pr(A \cap E_k) = \sum_{k=1}^{\infty} \Pr(A|E_k) \cdot \Pr(E_k)$$

Independence

- Events A and B are said to be independent if:

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

OR

$$\Pr(A|B) = \Pr(A) \quad \text{and} \quad \Pr(B|A) = \Pr(B)$$

- The collection of events $E_1, E_2, \ldots, E_n$, are independent if:

$$\Pr(E_1 \cap E_2 \cap \ldots \cap E_n) = \Pr(E_1) \cdot \Pr(E_2) \cdot \ldots \cdot \Pr(E_n)$$

- The events are mutually independent if for any subcollection $E_{i_1}, E_{i_2}, \ldots, E_{i_m}$, we have:

$$\Pr(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_m}) = \Pr(E_{i_1}) \cdot \Pr(E_{i_2}) \cdot \ldots \cdot \Pr(E_{i_m})$$

Bayes Theorem

$$\Pr(E_k|A) = \frac{\Pr(A|E_k)\cdot\Pr(E_k)}{\sum_{j=1}^{\infty}\Pr(A|E_j)\cdot\Pr(E_j)},$$

Example:

An insurance company classifies its policyholders according to three risk classes: L (low risk), M (medium risk) and H (high risk). The proportion of H policyholders is 20% and the proportion of L policyholders is 50%. For each of the risk classes, the probability of a claim is 0.01 for L, 0.02 for M, and 0.04 for H.

Q: If a claim occurs, what is the probability that it is from a Low Risk policyholder?

S: Let C = "the event that there is a claim", then:

$\Pr(L|C) = \frac{\Pr(L\cap C)}{\Pr(C)} = 26\%$, where
$\Pr(L\cap C) = \Pr(C|L)\cdot\Pr(L) = 0.01\cdot 0.5 = 0.005$
$\Pr(C) = \Pr(C|L)\cdot\Pr(L) + \Pr(C|M)\cdot\Pr(M) + \Pr(C|H)\cdot\Pr(H)$
$= 0.01\cdot 0.5 + 0.02\cdot 0.3 + 0.04\cdot 0.2 = 0.019$ (using LTP).

Counting Principles

- **Multiplication Rule:** Suppose $S_1, S_2, \ldots, S_M$ are *m* sets with respective number of elements $n_1, \ldots, n_M$. The number of wats of choosing one element from each set it given by:

  $$n_1 \cdot n_2 \cdot \ldots \cdot n_m.$$
- **Permutation:** The number of ways of arranging *n* distinct objects.
- **Combination:** The number of ways of choosing *r* objects from *n* distinct objects

  $$\binom{n}{r} \equiv \frac{n!}{r!\cdot(n-r)!}.$$
- **Multinomial:** The number of ways that *n* objects can be grouped into *r* classes with n_k in the *k*th class, where *k* = 1, 2, …, r, is given by:

  $$\binom{n}{n_1, n_2, \ldots, n_r} \equiv \frac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_r!}.$$

Counting Probabilities

Example:

Q: A drawer contains 10 pairs of socks. If 6 socks are taken at random without replacement, compute the probability that there is at least one matching pair among the 6 socks.

Solution: Let M = "at least one matching pair among 6 socks". We have $\Pr(M) = 1 - \Pr(M^c)$, where M^c = "no matching pair among 6 socks".
$\Pr(M^c) = 1\cdot\frac{18}{19}\cdot\frac{16}{18}\cdot\frac{14}{17}\cdot\frac{12}{16}\cdot\frac{10}{15} = 0.3467$. Hence,
$\Pr(M) = 1 - \Pr(M^c) = 1 - 0.3467 = 0.6533.$

Odds and Probabilities

- Odds: Ratio of the probability that the event will occur to the probability that it will not occur
  - E.g. if the odds are *a*:*b*, then the probability that the event will occur is:

$$\Pr(E) = \frac{a}{a+b}.$$

Random Variables and Distributions

- *Random Variable*: a quantity whose value depends on the outcome of a random experiment.
- *Cumulative Distribution Function* (cdf) is defined by:

$$F_X(x) = \Pr(X \leq x), \quad \text{for all } x.$$

- *Survival Function:*

$$S_X(x) = 1 - F_X(x) = \Pr(X > x).$$

Properties of a cumulative distribution function:

1. $F_X(\cdot)$ is a non-decreasing function, i.e., $F_X(x_1) \leq F_X(x_2)$ whenever $x_1 \leq x_2$;
2. $F_X(\cdot)$ is right-continuous, that is for all x, $\lim_{\epsilon \to 0^+} F_X(x+\epsilon) = F_X(x)$;
3. $F_X(-\infty) = 0$;
4. $F_X(+\infty) = 1$.

- Discrete random variable has a PMF (probability mass function):

$$p_X(x_k) = \Pr(X = x_k) = F_X(x_k) - F_X(x_{k-1})$$

- Continuous random variable has a PDF (probability density function):

$$f_X(x) = \frac{\partial}{\partial x} F_X(x)$$

Note that:

p.m.f. satisfies $\sum_{k=0}^{\infty} p_X(x_k) = 1$;
c.d.f. requires the right-continuous property;
p.d.f. satisfies $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

Measures of Location (Expected Value)

- The expected value of a RV *X* is:

$$\mu_X = \mathbb{E}[X] = \begin{cases} \sum_{\text{all } x} x \cdot p_X(x), \\ \text{if } X \text{ is discrete;} \\ \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx, \\ \text{if } X \text{ is continuous.} \end{cases}$$

- For a real valued function h(X):

$$\mathbb{E}[h(X)] = \begin{cases} \sum_{\text{all } x} h(x) \cdot p_X(x), \\ \text{if } X \text{ is discrete;} \\ \int_{-\infty}^{\infty} h(x) \cdot f_X(x)\, dx, \\ \text{if } X \text{ is continuous.} \end{cases}$$

Properties of the expected value operator:
Let X and Y be random variables, and $m, b \in \mathbb{R}$, we have:

$$\begin{aligned} \mathbb{E}[mX+b] = &\ m\mathbb{E}[X] + b; & \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]; \\ \mathbb{E}[X \cdot Y] = &\ \mathbb{E}[X] \cdot \mathbb{E}[Y], & \textit{only} \text{ if } X, Y \text{ independent.} \end{aligned}$$

Measures of Dispersion

- Mean Absolute Deviation (one measure of dispersion):

  $$MAD(X) = \mathbb{E}\left[|X - \mu_X|\right]$$

  - The MAD is minimised when μ_x is the median of the distribution
- **Variance** is the preferred measure of dispersion:

  $$\begin{aligned} \sigma^2 = Var(X) &= \mathbb{E}\left[(X-\mu_X)^2\right] \\ &= \mathbb{E}\left[X^2\right] - \mu_X^2. \end{aligned}$$

- Standard Deviation is given by:

  $$\sigma = \sqrt{Var(X)}.$$

Properties of the variance:
Let X and Y be random variables, and $m, b, c \in \mathbb{R}$, we have:

$$\begin{aligned} Var(c) = &\ 0; & Var(mX+b) = m^2 \cdot Var(X); \\ Var(X+Y) = &\ Var(X) + Var(Y), & \text{only if } X, Y \text{ independent.} \end{aligned}$$

Skewness

- The skewness of *X* is given by:

  $$\gamma_X = \mathbb{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^3\right]$$

- Coefficient of skewness:

  $$\alpha_3 = \frac{\mathbb{E}\left[X^3\right]}{\left(\mathbb{E}\left[X^2\right]\right)^{3/2}}.$$

- Properties:

  Properties of skewness (σ_{X+Y} is the s.d. of $X+Y$):

$$\gamma_X = 0, \quad \text{if the distribution of } X \text{ is symmetric;}$$
$$\gamma_{m \cdot X+b} = \text{sign}(m) \cdot \gamma_X;$$
$$\gamma_{X+Y} = \frac{\gamma_X \cdot \sigma_X^3 + \gamma_Y \cdot \sigma_Y^3}{\sigma_{X+Y}^3}, \quad \text{If } X, Y \text{ are independent.}$$

Kurtosis

- The (excess) kurtosis of X is given by:

$$\kappa_X = \mathbb{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^4\right] - 3.$$

  - It measures the peakedness or flatness of a distribution
- Kurtosis Coefficient:

$$\alpha_4 = \frac{\mathbb{E}\left[X^4\right]}{\left(\mathbb{E}\left[X^2\right]\right)^2}.$$

Moments

- The rth **central** moment of X is given by:

$$\mathbb{E}\left[(X-\mu_X)^r\right] = \begin{cases} \displaystyle\sum_{\text{all } x} (x-\mu_X)^r \cdot p_X(x), \\ \text{if } X \text{ is discrete;} \\ \displaystyle\int_{-\infty}^{\infty} (x-\mu_X)^r \cdot f_X(x)\, dx, \\ \text{if } X \text{ is continuous.} \end{cases}$$

- The rth **non-central** moment is given by:

$$\mathbb{E}\left[X^r\right] = \begin{cases} \displaystyle\sum_{\text{all } x} x^r \cdot p_X(x), \\ \text{if } X \text{ is discrete;} \\ \displaystyle\int_{-\infty}^{\infty} x^r \cdot f_X(x)\, dx, \\ \text{if } X \text{ is continuous.} \end{cases}$$

Moment Generating Function (MGF)

- The MGF of *X* is defined as:

$$\begin{aligned} M_X(t) &= \mathbb{E}\left[e^{X\cdot t}\right] \\ &\stackrel{*}{=} \mathbb{E}\left[1 + X\cdot t + X^2\cdot\frac{t^2}{2!} + X^3\cdot\frac{t^3}{3!} + \dots\right] \\ &= 1 + \mathbb{E}[X]\cdot t + \mathbb{E}\left[X^2\right]\cdot\frac{t^2}{2!} + \mathbb{E}\left[X^3\right]\cdot\frac{t^3}{3!} + \dots . \end{aligned}$$

Properties of m.g.f.:

$$\begin{aligned} M_{m\cdot X+b}(t) &= M_X(m\cdot t)\cdot e^{b\cdot t}, && \text{for constants } m, b; \\ M_{X+Y}(t) &= M_X(t)\cdot M_Y(t), && \text{only if } X, Y \text{ are independent.} \end{aligned}$$

- We can write the MGF as an infinite series of the moments as follows:

$$M_X(t) = \mathbb{E}\left[e^{X\cdot t}\right] = \sum_{k=0}^{\infty} \mu_k \cdot \frac{t^k}{k!}.$$

- We can generate the moments from the MGF using the relationship:

$$\mu_r = \mathbb{E}[X^r] = M_X^{(r)}(t)\Big|_{t=0}.$$

Probability Generating Function

- The PGF of an integer valued RV *Y* is:

$$P_Y(t) = \mathbb{E}\left[t^Y\right] = \sum_{i=0}^{\infty} p_Y(i)\cdot t^i.$$

Properties of p.g.f.:

- The relationship between p.g.f. and m.g.f. is as follows:

$$P_Y(t) = M_Y(\log(t)).$$

- Probabilities: $\Pr(Y=r) = P_Y^{(r)}(t)\Big|_{t=0} / r!$
- Take the k^{th} derivative and set $t=1$: $P_Y^{(k)}(1) = \mathbb{E}[Y\cdot(Y-1)\cdot(Y-2)\cdot\ldots\cdot(Y-k+1)]$.

Univariate Distributions

Bernoulli Distribution (Discrete)

- Outcome is one of two mutually exclusive events
    - Classified as a success (x = 1) or a failure (x = 0)

$$X = \begin{cases} 1, & \text{w.p. } p; \\ 0, & \text{w.p. } 1-p. \end{cases}$$

- **Notation**: $X \sim \text{Bernoulli}(p)$
- **Probability Mass Function:**

$$p_X(x) = p^x \cdot (1-p)^{1-x}$$

- **Expected Value:**

$$\begin{aligned}\mathbb{E}[X] &= \sum\nolimits_{\text{all } x} x \cdot p_X(x) \\ &= 0 \cdot (1-p) + 1 \cdot p = p.\end{aligned}$$

- **Variance:**

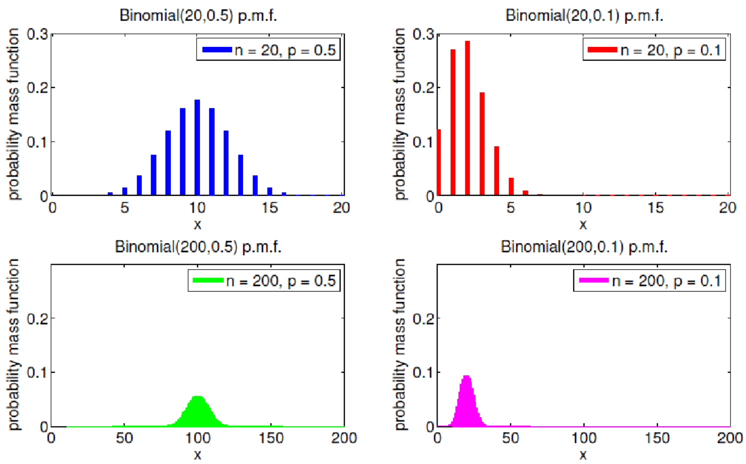
$$\begin{aligned}Var(X) &= \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 \\ &= \sum\nolimits_{\text{all } x} x^2 \cdot p_X(x) - p^2 \\ &= 0^2 \cdot (1-p) + 1^2 \cdot p - p^2 = p \cdot (1-p).\end{aligned}$$

- **Moment Generating Function:**

$$\begin{aligned}M_X(t) &= \mathbb{E}\left[e^{Xt}\right] \\ &= \sum\nolimits_{\text{all } x} e^{xt} \cdot p_X(x) \\ &= (1-p) \cdot e^{0 \cdot t} + pe^{1 \cdot t} = p \cdot e^t + (1-p).\end{aligned}$$

Binomial Distribution (Discrete)

- n independent trials of a Bernoulli RV
- X represents the number of successes out of n trials
- **Notation:** $X \sim \text{Binomial}(n, p)$



- **PMF:**

$$p_X(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

- **Expected Value:**

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathbb{E}[Y_i] \overset{*}{=} n \cdot \mathbb{E}[Y_i] = n \cdot p.$$

- **Variance:**

$$\begin{aligned} Var(X) &= Var\left(\sum_{i=1}^{n} Y_i\right) \\ &\overset{*}{=} \sum_{i=1}^{n} Var(Y_i) = n \cdot Var(Y_i) = n \cdot p \cdot (1-p). \end{aligned}$$

- **MGF**:

$$\begin{aligned} \mathbb{E}\left[e^{Xt}\right] &= \mathbb{E}\left[e^{\sum_{i=1}^{n} Y_i \cdot t}\right] \overset{*}{=} \prod_{i=1}^{n} \mathbb{E}\left[e^{Y_i \cdot t}\right] \overset{*}{=} \left[\mathbb{E}\left(e^{Y_i \cdot t}\right)\right]^n \\ &= \left(p \cdot e^t + (1-p)\right)^n \end{aligned}$$

Example:

Q: The probability of one or more accidents in a year for high risk insureds is 25%. When the insurer has 10H insureds, what's the probability that 5 or more insured have an accident?

Solution: X ="number of insureds with at least one accident a year".
$\Pr(X > 4) = 1 - \Pr(X \leq 4) = 1 - \sum_{i=0}^{4} \binom{10}{i} 0.25^i 0.75^{10-i} =$
$= 1 - 0.9219 = 0.0781.$

Geometric Distribution (Discrete)

- Consider a sequence of independent and repeated trials with p denoting the probability of success
- X is the number of trials required to obtain a first success
- **Notation:** X ~ Geometric(p)
- *Note:* Memoryless Property, i.e.
  $\Pr(X \leq a | X > b) = \Pr(X \leq a - b)$
- **PMF:**

$$p_X(x) = p \cdot (1-p)^{x-1}$$

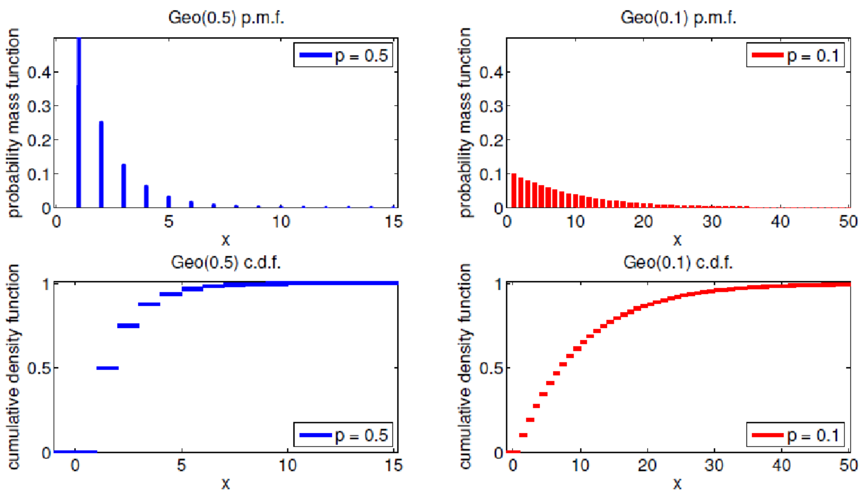
- **Expected Value:**

$$\mathbb{E}[X] = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

- **Variance:**
  $Var(X) = \frac{1-p}{p^2}$

- **MGF:**

$$
\begin{aligned}
M_X(t) =& \mathbb{E}\left[e^{X\cdot t}\right] = \sum_{x=1}^{\infty} e^{x\cdot t} \cdot p \cdot (1-p)^{x-1} \\
=& p \cdot e^t \cdot \sum_{z=0}^{\infty} e^{t\cdot z} \cdot (1-p)^z \\
=& p \cdot e^t \cdot \sum_{z=0}^{\infty} \left(e^t \cdot (1-p)\right)^z \\
=& \frac{p \cdot e^t}{1-(1-p)\cdot e^t}.
\end{aligned}
$$



Negative Binomial Distribution (Discrete)

- *X* is the random variable that denotes the number of trials required until there are *r* successes.
- **Notation:** *X* ~ N.B.(r,p)
- **PMF:**

$$
p_X(x) = \binom{x-1}{r-1} \cdot p^r \cdot (1-p)^{x-r},
$$

- **Expected Value:**

$$
\mathbb{E}[X] = \mathbb{E}[Y_1 + \ldots + Y_r] = r \cdot \mathbb{E}[Y_i] = \frac{r}{p},
$$

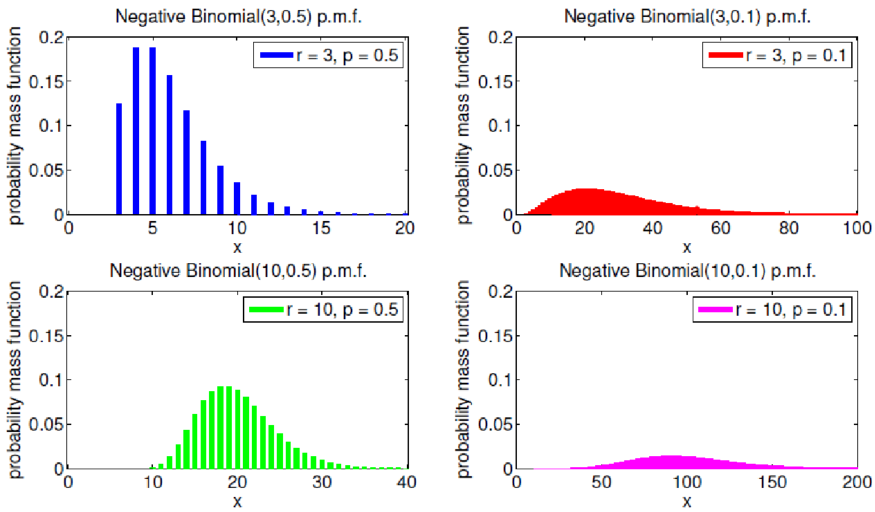
- **Variance:**

$$
Var(X) = Var(Y_1 + \ldots + Y_r) = r \cdot Var(Y_i) = \frac{r \cdot (1-p)}{p^2}.
$$

- MGF:

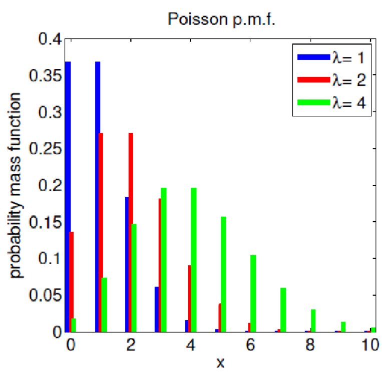
$$
M_X(t) = M_{Y_1+\ldots+Y_r}(t) = (M_{Y_i}(t))^r = \left(\frac{p \cdot e^t}{1-(1-p)\cdot e^t}\right)^r \tag{8}
$$

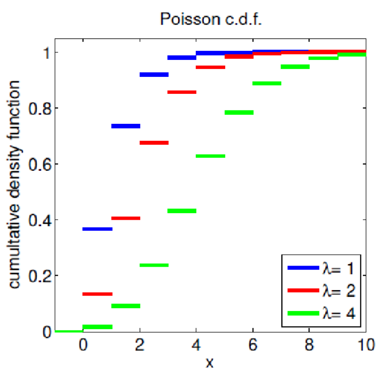
- Examples:



Poisson Distribution (Discrete)

- *X* is a random variable which represents the number of rare events that occur in a time period
- **Notation:** *X* ~ Poisson(λ)





- **PMF:**

$$p_X(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!},$$

- **Expected Value:**

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{\text{all } x} x \cdot p_X(x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x \cdot e^{-\lambda}}{x!} \\
&\overset{*}{=} \lambda \cdot \sum_{x=1}^{\infty} \frac{\lambda^{x-1} \cdot e^{-\lambda}}{(x-1)!} \\
&\overset{**}{=} \lambda \cdot \sum_{z=0}^{\infty} \frac{\lambda^z \cdot e^{-\lambda}}{z!} \overset{***}{=} \lambda.
\end{aligned}$$

- **Variance**

$$\mathbb{E}\left[X^2\right] = \sum_{\text{all } x} x^2 \cdot p_X(x) = \sum_{x=0}^{\infty} (x \cdot (x-1) + x) \cdot \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$
$$= \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x \cdot e^{-\lambda}}{x!} + \sum_{x=0}^{\infty} (x \cdot (x-1)) \cdot \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$
$$= \lambda + \lambda^2 \cdot \sum_{x=2}^{\infty} \frac{\lambda^{x-2} \cdot e^{-\lambda}}{(x-2)!} \stackrel{**}{=} \lambda + \lambda^2 \cdot \sum_{z=0}^{\infty} \frac{\lambda^z \cdot e^{-\lambda}}{z!}$$
$$\stackrel{***}{=} \lambda + \lambda^2;$$

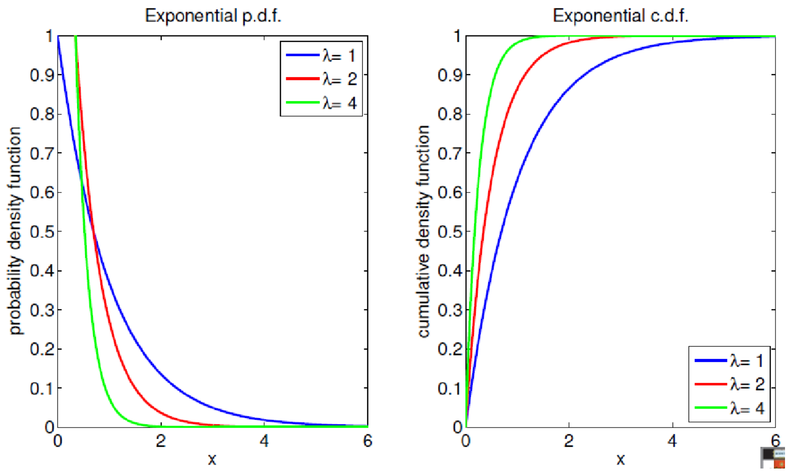
- **MGF:**

$$M_X(t) = \exp\left(\lambda\left(e^t - 1\right)\right)$$

Exponential Distribution (Continuous)

- Notation: $X \sim \text{EXP}(\lambda)$
- PDF:

$$f_X(x) = \lambda \cdot e^{-\lambda \cdot x}, \quad \text{for } x \geq 0.$$



Mean: $\mathbb{E}[X] = \frac{1}{\lambda}$, **Variance**: $Var(X) = \frac{1}{\lambda^2}$,
and **m.g.f**: $M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}$.

- Memoryless Property:

$$\Pr(X > a + b | X > a) = \frac{\Pr(X > a + b, X > a)}{\Pr(X > a)}$$
$$= \frac{\Pr(X > a + b)}{\Pr(X > a)}$$
$$= \frac{\int_{a+b}^{\infty} \lambda \cdot e^{-\lambda x} dx}{\int_{a}^{\infty} \lambda \cdot e^{-\lambda x} dx}$$
$$= \frac{e^{-\lambda \cdot (a+b)}}{e^{-\lambda \cdot a}} = e^{-\lambda \cdot b}$$
$$= \Pr(X > b).$$

Gamma Distribution (Continuous)

- Typically used for insurance claim modelling
- Notation: $X \sim \text{Gamma}(\alpha,\beta)$
- *Special* Case: If $\alpha = 1$, we have an exponential distribution with parameter β
- PDF:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta \cdot x},$$

Mean: $\mathbb{E}[X] = \frac{\alpha}{\beta}$; variance: $Var(X) = \frac{\alpha}{\beta^2}$;

- MGF:

$$M_X(t) = \mathbb{E}\left[e^{Xt}\right] = \left(1 - \frac{t}{\beta}\right)^{-\alpha}, \; t < \beta$$

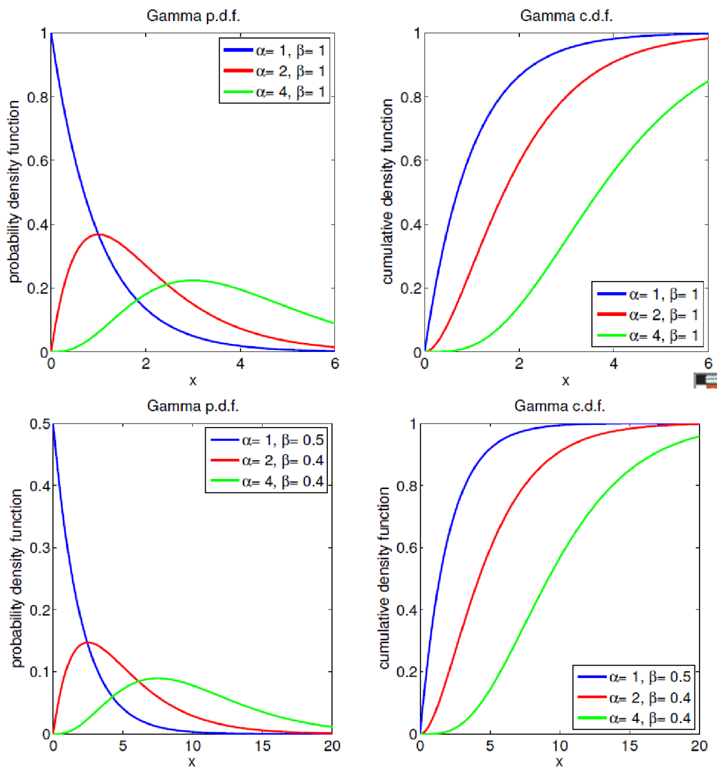
- The Gamma Function

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

  - Recursive Relationships:

$$\begin{aligned} \Gamma(\alpha+1) &= \alpha \cdot \Gamma(\alpha) \\ \Gamma(n) &= (n-1)!, \end{aligned}$$

- Examples



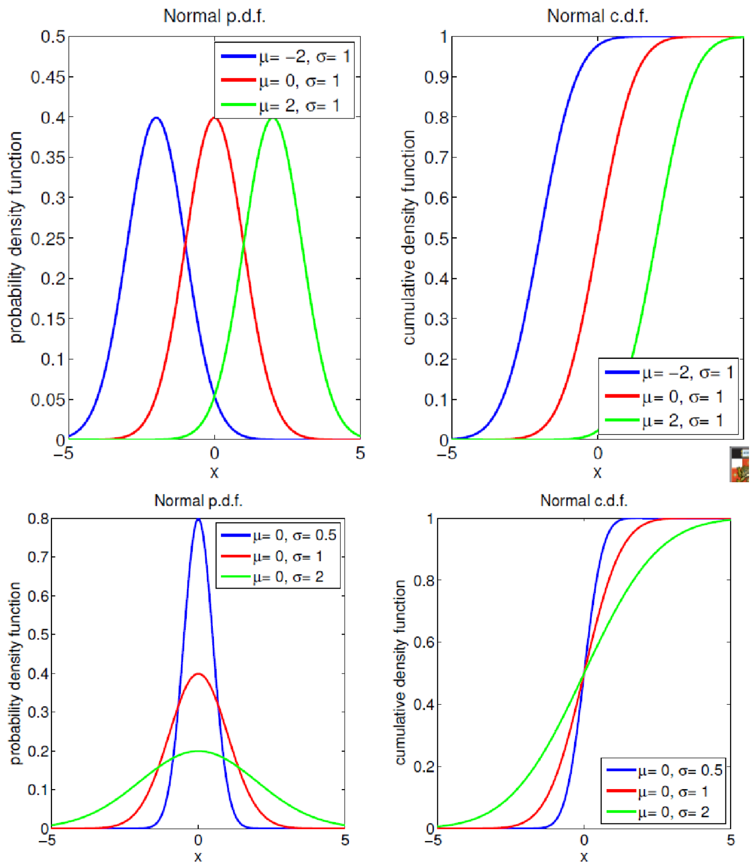
The Normal (Gaussian) Distribution (Continuous)

- **Notation**: $X \sim N(\mu,\sigma^2)$
  - μ - Mean
  - σ^2 – Variance
- PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\exp\left(-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

- MGF:

$$M_X(t) = \mathbb{E}\left[e^{Xt}\right] = \exp\left(\mu\cdot t + \frac{1}{2}\cdot\sigma^2\cdot t^2\right)$$



- **Standard Normal**:
  - $\mu = 0$
  - $\sigma^2 = 1$
  - We can standardise a normal random variable $X \sim N(\mu,\sigma^2)$ as follows:

$$Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1), \quad \text{and} \quad X = \sigma Z + \mu.$$

Lognormal Distribution (Continuous)

- **Notation:** $X \sim \text{Lognormal}(\mu,\sigma^2)$

  If $Y \sim \mathrm{N}(\mu,\sigma^2)$ and $X = e^Y$ then X is lognormal; or $\log(X) \sim \mathrm{N}(\mu,\sigma^2)$, i.e., "log is normally distributed".

  $X = \exp(Y) \sim \text{Lognormal}(\mu,\sigma^2)$ is said to have a lognormal distribution with parameters μ and σ^2.

- PDF:

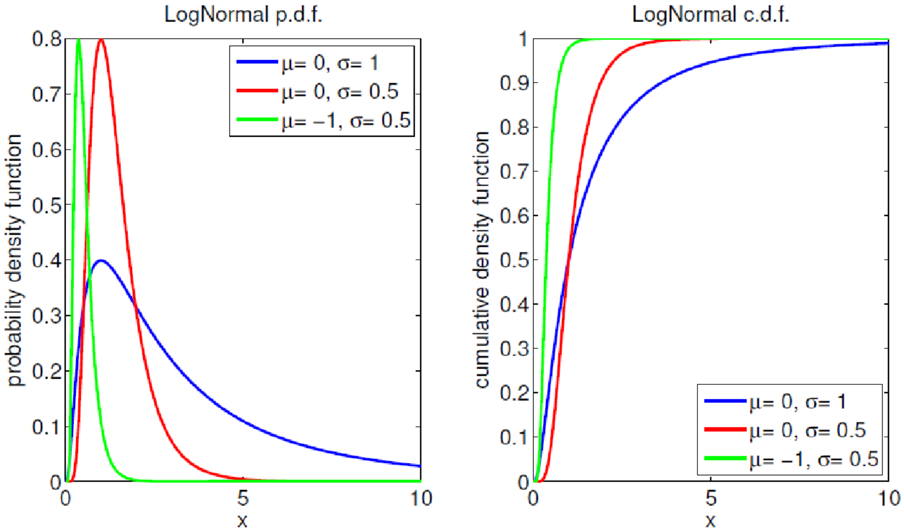
$$f_X(x) = \frac{1}{x \cdot \sigma \cdot \sqrt{2\pi}} \exp\left(-\frac{1}{2} \cdot \left(\frac{\log(x) - \mu}{\sigma}\right)^2\right),$$

- Expected Value:

$$\mathbb{E}[X] = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

- Variance:

$$Var(X) = e^{\left(2\mu+\sigma^2\right)} \cdot \left(e^{\sigma^2} - 1\right).$$



- Applications:
  - Models of stock prices are often based on the lognormal distribution
  - Also used to model claim sizes
- Properties:
  - Product of independent lognormal random variables are lognormal
  - To calculate probabilities for a lognormal random variable, restate them as possibilities about the associated normal random variable

$$\begin{aligned}\Pr(X \le a) &= \Pr(\log(X) \le \log(a)) \\ &= \Pr\left(\frac{\log(X) - \mu}{\sigma} \le \frac{\log(a) - \mu}{\sigma}\right) \\ &= \Pr\left(Z \le \frac{\log(a) - \mu}{\sigma}\right).\end{aligned}$$

Uniform Distribution (Continuous)

- An experiment where all outcomes are equally likely to happen
- **Notation**: $X \sim$ UNIF(a,b)
- PDF:

$$f_X(x) = \frac{1}{(b-a)},$$

- CDF:

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx = \begin{cases} \int_{-\infty}^{x} 0\,dx & = & 0, & \text{if } x < a; \\ \int_{a}^{x} \frac{1}{(b-a)}dx & = & \frac{x-a}{b-a}, & \text{if } a \le x \le b; \\ \int_{a}^{b} \frac{1}{(b-a)}dx & = & 1, & \text{if } x > b. \end{cases}$$

- Expected Value:

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x)dx \\ &= \int_{a}^{b} x \cdot \frac{1}{(b-a)}dx = \left[\frac{1}{2}\frac{x^2}{b-a}\right]_a^b = \frac{a+b}{2}.\end{aligned}$$

- Variance:

$$\begin{aligned}Var(X) =& \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 \\ =& \int_{a}^{b} x^2 \frac{1}{(b-a)}dx - \left(\frac{a+b}{2}\right)^2 \\ =& \left[\frac{1}{3}\frac{x^3}{b-a}\right]_a^b - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}\end{aligned}$$

- MGF:

$$\begin{aligned}M_X(t) =& \mathbb{E}\left[e^{Xt}\right] \\ =& \int_{-\infty}^{\infty} e^{xt} \cdot f_X(x)dx \\ =& \int_{a}^{b} e^{xt} \cdot \frac{1}{(b-a)}dx \\ =& \frac{1}{(b-a)}\left[\frac{e^{xt}}{t}\right]_a^b = \frac{e^{bt}-e^{at}}{t(b-a)}.\end{aligned}$$

Beta Distribution

- Generally used to model proportions because the range is from of x is between 0 and 1.
- **Notation**: X ~ Beta(a,b)
  - a > 0 (shape parameter)
  - b > 0 (shape parameter)
- Beta function:

$$\begin{aligned}B(\alpha,\beta) &= \int_{0}^{1} x^{\alpha-1} \cdot (1-x)^{\beta-1}dx \\ &= \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}.\end{aligned}$$

- PDF:

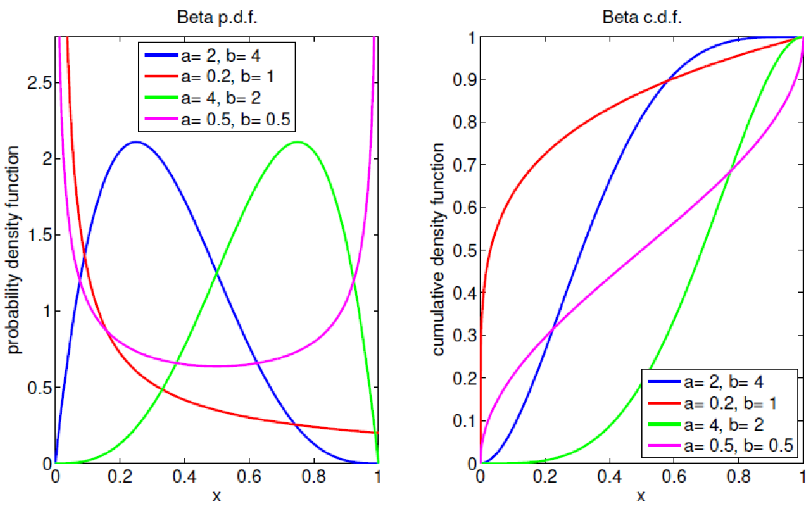
$$f_X(x) = \underbrace{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}}_{1/B(a,b)} x^{a-1}(1-x)^{b-1},$$

- Expected Value:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx = \int_0^1 x \cdot \left(\frac{x^{a-1} \cdot (1-x)^{b-1}}{B(a,b)}\right) dx$$
$$= \frac{1}{B(a,b)} \cdot \int_0^1 x^a \cdot (1-x)^{b-1} dx$$
$$= \frac{B(a+1,b)}{B(a,b)}$$
$$= \frac{\Gamma(a+b)}{\Gamma(a) \cdot \Gamma(b)} \cdot \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}$$
$$= \frac{\Gamma(a+b)}{\Gamma(a) \cdot \Gamma(b)} \cdot \frac{a\Gamma(a) \cdot \Gamma(b)}{(a+b)\Gamma(a+b)} = \frac{a}{a+b}$$

- Variance:

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{a \cdot b}{(a+b)^2(a+b+1)}.$$



Pareto Distribution (Continuous)

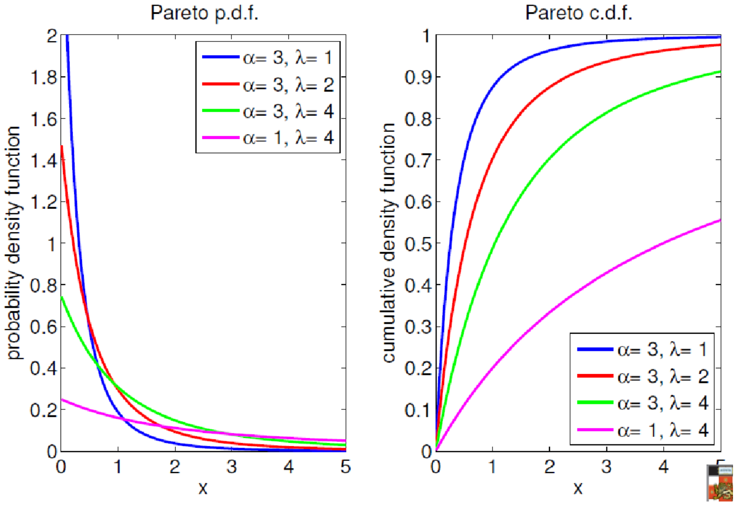
- Heavy tailed distribution
- Often used for reinsurance purposes. It tapers away to zero more slowly than lognormal, hence it is more appropriate for estimating reinsurance in respect for very large claims.
- CDF:

$$F_X(x) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^{\alpha}, \quad x > 0.$$

- PDF:

$$f_X(x) = \frac{\alpha \cdot \lambda^{\alpha}}{(\lambda + x)^{\alpha+1}} = \frac{\alpha}{\lambda \cdot (1 + x/\lambda)^{\alpha+1}}.$$

- Examples



- Moments (Do not always exist):

$$\mathbb{E}\left[X^r\right] = \frac{\Gamma\left(\alpha - r\right) \cdot \Gamma\left(1 + r\right)}{\Gamma\left(\alpha\right)} \cdot \lambda^r,$$

  for $r = 1, 2$ and 3, provided $r < \alpha$.

  - E.g.

$$Var\left(X\right) = \frac{\alpha \cdot \lambda^2}{\left(\alpha - 1\right)^2 \cdot \left(\alpha - 2\right)}$$

    and only exists if $\alpha > 2$.

Joint and Multivariate Distribution

The Bivariate Case

- Consider a pair of random variables, (*X, Y*). The joint distribution function is:

$$F_{X,Y}(x, y) = \Pr\left(X \leq x, Y \leq y\right)$$

- For **discrete** random variables:

$$p_{X,Y}\left(x_i, y_j\right) = \Pr\left(X = x_i, Y = x_j\right)$$

- And the marginal pmf of X and Y respectively:

$$p_X\left(x_i\right) = \sum_{j=1}^{\infty} p_{X,Y}\left(x_i, y_j\right)$$

$$p_Y\left(y_j\right) = \sum_{i=1}^{\infty} p_{X,Y}\left(x_i, y_j\right)$$

- For **continuous** random variables:

$$f_{X,Y}\left(x, y\right) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{X,Y}\left(x, y\right)$$

$$F_{X,Y}\left(x, y\right) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}\left(u, v\right) du dv.$$

- The marginal density of X and Y respectively:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx.$$

- The marginal CDF of X and Y respectively:

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du \qquad \text{and} \qquad F_Y(y) = \int_{-\infty}^{y} f_Y(u)\, du,$$

  - Alternatively:

$$F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(u, v)\, dvdu$$

$$\text{and } F_Y(y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f_{X,Y}(u, v)\, dudv.$$

Means

- The of a vector of X and Y, $\underline{V}$, is the vector whose elements are the means of X and Y:

$$\mathbb{E}[\underline{V}] = \begin{bmatrix} \mathbb{E}[X] \\ \mathbb{E}[Y] \end{bmatrix} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}$$

If $X_1, X_2, \ldots, X_n$ are random variables with expectations $\mathbb{E}[X_i]$ for $i = 1, \ldots, n$ and Y is a affine function of the X_i, i.e.,

$$Y = a + \sum_{i=1}^{n} b_i X_i, \quad \text{then, we have the additively rule:}$$

$$\mathbb{E}[Y] = \mathbb{E}\left[a + \sum_{i=1}^{n} b_i X_i\right] = \quad a + \sum_{i=1}^{n} \mathbb{E}[b_i X_i] = a + \sum_{i=1}^{n} b_i \mathbb{E}[X_i].$$

Variance and Covariance

- The variance of $\underline{V}$ is:

$$Var(\underline{V}) = \begin{bmatrix} Var(X) & Cov(X, Y) \\ Cov(X, Y) & Var(Y) \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

- And the covariance is:

$$\begin{aligned} Cov(X, Y) \equiv \sigma_{XY} &= \mathbb{E}[(X - \mu_X) \cdot (Y - \mu_Y)] \\ &= \mathbb{E}[X \cdot Y - X \cdot \mu_Y - \mu_X \cdot Y + \mu_X \cdot \mu_Y] \\ &= \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y]. \end{aligned}$$

- Properties of covariance:
  - If X and Y are jointly distributed:

$$\begin{aligned} Cov(X, Y) &= \mathbb{E}[(X - \mu_X) \cdot (Y - \mu_Y)] \\ &= \mathbb{E}[X \cdot Y - X \cdot \mu_Y - Y \cdot \mu_X + \mu_X \cdot \mu_Y] \\ &= \mathbb{E}[X \cdot Y] - \mu_X \cdot \mu_Y. \end{aligned}$$

  - If X and Y are independent

$$Cov(X, Y) = \mathbb{E}[X \cdot Y] - \mu_X \cdot \mu_Y \stackrel{*}{=} \mathbb{E}[X] \cdot \mathbb{E}[Y] - \mu_X \cdot \mu_Y = 0$$

- For RV's *X, Y, Z*:

$$
\begin{aligned}
Cov\,(a + X, Y) =& \mathbb{E}\left[(a + X - (a + \mu_X)) \cdot (Y - \mu_Y)\right] \\
=& \mathbb{E}\left[(X - \mu_X) \cdot (Y - \mu_Y)\right] \\
=& Cov\,(X, Y) \\
Cov\,(a \cdot X, b \cdot Y) =& \mathbb{E}\left[(a \cdot X - a \cdot \mu_X) \cdot (b \cdot Y - b \cdot \mu_Y)\right] \\
=& \mathbb{E}\left[a \cdot (X - \mu_X) \cdot b \cdot (Y - \mu_Y)\right] \\
=& a \cdot b \cdot \mathbb{E}\left[(X - \mu_X) \cdot (Y - \mu_Y)\right] = a \cdot b \cdot Cov\,(X, Y) \\
Cov\,(X, Y + Z) =& \mathbb{E}\left[(X - \mu_X) \cdot (Y + Z - \mu_Y - \mu_Z)\right] \\
=& \mathbb{E}\left[(X - \mu_X) \cdot ((Y - \mu_Y) + (Z - \mu_Z))\right] \\
=& \mathbb{E}\left[(X - \mu_X) \cdot (Y - \mu_Y) + (X - \mu_X) \cdot (Z - \mu_Z)\right] \\
=& Cov\,(X, Y) + Cov\,(X, Z)\,.
\end{aligned}
$$

- Suppose:

$$U = a + \sum_{i=1}^{n} b_i \cdot X_i \quad \text{and} \quad V = c + \sum_{j=1}^{m} d_j \cdot Y_j.$$

  Then:

$$Cov\,(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i \cdot d_j \cdot Cov\,(X_i, Y_j)$$

- Variance of *(X + Y)* and *(aX)*:

$$
\begin{aligned}
Var\,(X + Y) =& Cov(X + Y, X + Y) \\
=& Cov(X, X) + Cov(Y, Y) + 2Cov(X, Y) \\
=& Var\,(X) + Var\,(Y) + 2Cov\,(X, Y)\,.
\end{aligned}
$$

$$Var\,(aX) = Cov\,(aX, aX) = a^2\,Cov\,(X, X) = a^2\,Var\,(X)$$

Correlation Coefficient

- Correlation coefficient:

$$\rho\,(X, Y) = \frac{Cov\,(X, Y)}{\sqrt{Var\,(X) \cdot Var\,(Y)}},$$

- Note that:

$$-1 \leq \rho\,(X, Y) \leq 1.$$

Question: Does a correlation of zero imply independence?

Solution: Note that we have that if X, Y are independent, then $Cov(X, Y) = 0$, hence:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}} = \frac{0}{\sqrt{Var(X) \cdot Var(Y)}} = 0.$$

However, the reverse does not need to hold.

Conditional Distributions

- Let X,Y be discrete RVs with joint PMF. The conditional probability of X given Y is:

$$\Pr(X = x_i | Y = y_j) = \frac{\Pr(X = x_i, Y = y_j)}{\Pr(Y = y_j)}.$$

- Let X, Y be continuous RVs with joint PMF. The conditional density of Y given X is:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

The Bivariate Normal Distribution

- Suppose $[X,Y]^T$ has a bivariate normal distribution, then its density is given by:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}A\right)$$

  - Where:

$$A = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\cdot\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2.$$

- These results follow:
  1. The marginals are: $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$.
  2. The conditional distributions are:

$$(Y|X = x) \sim N\left(\mu_Y + \rho(x-\mu_X)\frac{\sigma_Y}{\sigma_X}, \sigma_Y^2(1-\rho^2)\right)$$

  and

$$(X|Y = y) \sim N\left(\mu_X + \rho(y-\mu_Y)\frac{\sigma_X}{\sigma_Y}, \sigma_X^2(1-\rho^2)\right).$$

  3. The correlation coefficient between X and Y is: $\rho(X,Y) = \rho$.

Laws

- The law of iterated expectations:

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$$

- The conditional variance identity:

$$Var(Y) = Var(\mathbb{E}[Y|X]) + \mathbb{E}[Var(Y|X)]$$

- Application:

Mean of S: The mean of the aggregate claims is:

$$\mathbb{E}[S] = \mathbb{E}[X_i] \cdot \mathbb{E}[N].$$

th the
m filled.
e X_i and N
iance of a

This is straightforward:

$$\begin{aligned}
\mathbb{E}[S] =& \mathbb{E}[\mathbb{E}[S\,|N]] \\
=& \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N} X_i\,|N\right]\right] \\
=& \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{E}[X_i|N]\right] \\
\stackrel{*}{=}& \mathbb{E}[N \cdot \mathbb{E}[X_i]] \\
=& \mathbb{E}[\mathbb{E}[X_i]] \cdot \mathbb{E}[N] = \mathbb{E}[X_i] \cdot \mathbb{E}[N].
\end{aligned}$$

Variance of S: The variance of the aggregate claims is:

$$Var(S) = (\mathbb{E}[X_i])^2 \cdot Var(N) + \mathbb{E}[N] \cdot Var(X_i).$$

This is also straightforward to show:

$$\begin{aligned}
Var(S) \stackrel{*}{=}& \mathbb{E}[Var(S\,|N)] + Var(\mathbb{E}[S\,|N]) \\
=& \mathbb{E}\left[Var\left(\sum_{i=1}^{N} X_i\middle|N\right)\right] + Var(\mathbb{E}[X_i] \cdot N) \\
\stackrel{**}{=}& \mathbb{E}[N] \cdot \mathbb{E}\left[\underbrace{Var(X_i)}_{\text{constant}}\right] + \left(\underbrace{\mathbb{E}[X_i]}_{\text{constant}}\right)^2 \cdot Var(N) \\
=& \mathbb{E}[N] \cdot Var(X_i) + (\mathbb{E}[X_i])^2 \cdot Var(N)
\end{aligned}$$

The Multivariate Case

- Let $\underline{X} = [X_1, X_2, \ldots, X_n]^T$ be a random vector with n elements. The joint distribution of $\underline{X}$ is denoted by:
  $$F_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n) = \Pr(X_1 \le x_1, \ldots, X_n \le x_n)$$
- In the discrete case:
  $$p_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n) = \Pr(X_1 = x_1, \ldots, X_n = x_n)$$
- In the continuous case:
  $$f_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n) = \frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n} F_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n)$$
  - The joint DF is given by:
    $$F_{X_1,X_2,\ldots,X_n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1,X_2,\ldots,X_n}(z_1, \ldots, z_n)\, dz_1 \ldots dz_n.$$
  - To derive marginal pdfs, evaluate over the region except for the variable of interest:
    $$f_{X_k}(x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,X_2,\ldots,X_n}(z_1, \ldots, x_k \ldots, z_n) \prod_{j \neq k} dz_j.$$

Independent Random Variables

- The random variables $X_1, X_2, \dots, X_n$ are said to be independent if their joint density can be written as:

  $$F_{X_1,X_2,\dots,X_n}(x_1,\dots,x_n) = F_{X_1}(x_1)\cdot\dots\cdot F_{X_n}(x_n)$$

  As a consequence, their joint density can also be written as:

  $$f_{X_1,X_2,\dots,X_n}(x_1,\dots,x_n) = f_X(x_1)\cdot\dots\cdot f_{X_n}(x_n),$$

  in the continuous case and for the discrete case as:

  $$p_{X_1,X_2,\dots,X_n}(x_1,\dots,x_n) = p_X(x_1)\cdot\dots\cdot p_{X_n}(x_n).$$

- Further:

  Also, we have (if independent):

  $$\mathbb{E}[X_1\cdot X_2\cdot\dots\cdot X_n] = \mathbb{E}[X_1]\cdot\mathbb{E}[X_2]\cdot\dots\cdot\mathbb{E}[X_n],$$

  and in fact for *any* function $g_1,\dots,g_n$, (if independent) we have:

  $$\mathbb{E}[g_1(X_1)\cdot g_2(X_2)\cdot\dots\cdot g_n(X_n)] = \mathbb{E}[g_1(X_1)]\cdot\mathbb{E}[g_2(X_2)] \cdot\dots\cdot\mathbb{E}[g_n(X_n)].$$

- If X and Y are independent, then:

  1. $Cov[X,Y] = 0$ and so $\rho(X,Y) = 0$.
  2. $\mathbb{E}[X\,|\,Y] = \mathbb{E}[X]$ and of course $\mathbb{E}[Y\,|\,X] = \mathbb{E}[Y]$.
  3. A very useful result about independence is that $X_1, X_2, \dots, X_n$ are independent if and only if we can write the joint distribution as a product of functions that involve only each random variable:

     $$F_{X_1,X_2,\dots,X_n}(x_1,\dots,x_n) = H_{X_1}(x_1)\cdot\dots\cdot H_{X_n}(x_n)$$

     for some functions $H_{X_1},\dots,H_{X_n}$.

Sampling and Summarizing Data

- Population vs Sample:
  - Population: the large body of data
  - Sample: a subset of the population

Summarising Data: Numerical Approaches

- Given a set of observations $x_1, x_2, \dots, x_n$ selected from a population:

  Sorted data in ascending order: $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, such that $x_{(1)}$ is the smallest and $x_{(n)}$ is the largest.

Measures of Location

- Sample mean:

$$\overline{x} = \frac{1}{n} \cdot \sum_{k=1}^{n} x_k$$

- The population mean:

$$\mu_X = \sum_{\text{all } x} p_X(x) \cdot x$$

- The 100α% trimmed mean (the average of observations after discarding the lowest 100α% and the highest 100α%):

$$\widetilde{x}_\alpha = \frac{x_{(\lfloor n\alpha \rfloor+1)} + \dots + x_{(n-\lfloor n\alpha \rfloor)}}{n - 2\lfloor n\alpha \rfloor}$$

Measures of Spread

- Sample variance:

$$\begin{aligned} s^2 &= \frac{1}{n-1} \cdot \sum_{k=1}^{n} (x_k - \overline{x})^2 = \frac{1}{n-1} \cdot \left(\sum_{k=1}^{n} x_k^2 + \sum_{k=1}^{n} \overline{x}^2 - 2 \sum_{k=1}^{n} x_k \overline{x} \right) \\ &= \frac{1}{n-1} \cdot \left(\sum_{k=1}^{n} x_k^2 - n \cdot \overline{x}^2 \right). \end{aligned}$$

- The population variance:

$$\sigma^2 = Var(X) = \sum_{\text{all } x} p_X(x) \cdot (x - \mu_X)^2 = \sum_{\text{all } x} p_X(x) \cdot x^2 - \mu_X^2$$

Quantiles

Mode: The mode m is the value that maximises the p.m.f. $p_X(x)$ in the discrete case or the p.d.f. $f_X(x)$ in the continuous case.

Median, M:

$$M = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{if } n \text{ is odd;} \\ \frac{1}{2} \cdot \left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right), & \text{if } n \text{ is even.} \end{cases}$$

Median absolute deviation:

$$MAD = \text{median of the numbers:} \{|x_i - M|\}.$$

Range:

$$R = x_{(n)} - x_{(1)}.$$

Interquartile range:

$$IQR = Q_3 - Q_1.$$

Empirical CDF

- Given a set of observations $x_1, x_2, \dots, x_n$ the ECDF is given by:

$$F_n(x) = \frac{1}{n} \cdot (\text{number of observations } \leq x)$$

$$\mathbb{E}[F_n(x)] = F_X(x)$$

$$Var(F_n(x)) = \frac{1}{n} \cdot F_X(x) \cdot (1 - F_X(x)).$$

Properties of Sample Mean and Sample Variance

- Suppose you select randomly a sample from a much larger population
- The standard assumption is that this sample represents *n* i.i.d random variables.
- The sample mean is:

$$\overline{X} = \frac{1}{n} \cdot \sum_{k=1}^{n} X_k,$$

- The sample variance is:

$$S^2 = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (X_k - \overline{X})^2.$$

- We use capital letters to denote a random variable coming from a certain distribution. We use lower case to denote an observation from this RV.
- Hence, when we write $\underline{X}$, we mean the random variable made of all other random variables $X_1, X_2, \dots, X_n$:

$$\bar{X} = \sum X_i / n.$$

- Expected value of the sample mean:

$$\begin{aligned}
\mathbb{E}[\overline{X}] &= \mathbb{E}\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] \\
&= \frac{\sum_{i=1}^{n} \mathbb{E}[X_i]}{n} \\
&= \frac{n \cdot \mu}{n} = \boldsymbol{\mu}.
\end{aligned}$$

- Variance of the sample mean:

$$\begin{aligned}
Var(\overline{X}) &= Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) \\
&= \frac{\sum_{i=1}^{n} Var(X_i)}{n^2} \\
&= \frac{n \cdot \sigma^2}{n^2} \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

Order Statistics

- $X_1, X_2, \ldots, X_n$ are n i.i.d Rvs with common distribution. Sort these variables and denote by:

  $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$

  In particular, $X_{(1)} = \min\{X_1, \ldots, X_n\}$ is the minimum and $X_{(n)} = \max\{X_1, \ldots, X_n\}$ is the maximum.

  For simplicity, denote $U = X_{(n)}$ and $V = X_{(1)}$.
- Distribution of the **Maximum**:

$$\begin{aligned} F_U(u) &= \Pr(U \leq u) \\ &\overset{*}{=} \Pr(X_1 \leq u) \cdot \Pr(X_2 \leq u) \cdot \ldots \cdot \Pr(X_n \leq u) \\ &= (F_X(u))^n, \end{aligned}$$

- Density of the Maximum:

$$f_U(u) = n \cdot f_X(u) \cdot (F_X(u))^{n-1}$$

- Distribution of the **Minimum**:

$$\begin{aligned} F_V(v) &= \Pr(V \leq v) = 1 - \Pr(V > v) \\ &\overset{*}{=} 1 - (\Pr(X_1 > v) \cdot \ldots \cdot \Pr(X_n > v)) \\ &= 1 - (1 - F_X(v))^n, \end{aligned}$$

- Density of the Maximum:

$$f_V(v) = n \cdot f_X(v) \cdot (1 - F_X(v))^{n-1}$$

- The joint density of order statistics is given by:

$$f_{X_{(1)}, X_{(2)}, \ldots, X_{(n)}}(y_1, y_2, \ldots, y_n) = n! \cdot f_X(y_1) \cdot f_X(y_2) \cdot \ldots \cdot f_X(y_n).$$

- Distribution of the kth order statistic:

  Using heuristics, the probability density of the k^{th} order statistic is given by: $f_K(x) =$

$$\underbrace{\frac{n!}{(k-1)!\,(n-k)!}}_{\#\text{ possible ordering}} \cdot \underbrace{f_X(x)}_{\substack{\text{1 observa-}\\\text{tion equals } x}} \cdot \underbrace{(F_X(x))^{k-1}}_{\substack{k-1\text{ observa-}\\\text{tions smaller}}} \cdot \underbrace{(1 - F_X(x))^{n-k}}_{\substack{n-k\text{ observa-}\\\text{tions larger}}}.$$

Example:

- Consider an insurance company with n branches. Assume that the lifetimes of the branches are $T_1, T_2, \ldots, T_n$ which are i.i.d with exponential distribution with parameter λ.
- Suppose that the insurance company will go bankrupt if any one of the branches goes bankrupt. The lifetime V of the insurance company is therefore the minimum of the T_K:

  $V = \min\{T_1, \ldots, T_n\}$
- The density of V is:

$$\begin{aligned} f_V(v) &= n \cdot f_T(v) \cdot (1 - F_T(v))^{n-1} \\ &= n \cdot \lambda \cdot e^{-\lambda \cdot v} \cdot \left(e^{-\lambda \cdot v}\right)^{n-1} = (n \cdot \lambda) \cdot e^{-(n \cdot \lambda) \cdot v} \end{aligned}$$

- Suppose the branches are connected in parallel, that is, the insurance company will go bankrupt only if all the branches go bankrupt. The lifetime *U* of the system is therefore:

$$U = \max\{T_1, \ldots, T_n\}$$

- The density of *U* is:

$$\begin{aligned} f_U(u) &= n \cdot f_T(u) \cdot (F_T(u))^{n-1} \\ &= n \cdot \lambda \cdot e^{-\lambda \cdot u} \cdot \left(1 - e^{-\lambda \cdot u}\right)^{n-1} \end{aligned}$$

Functions of Random Variables

The CDF Technique

- Let *X* be a continuous random variable.
- Suppose that *Y = g(X)* is a function of *X*.
- Suppose g() is differentiable and strictly **increasing**:
  - ▶ The CDF of Y can be derived using:

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr(g(X) \leq y) \\ &= \Pr\left(X \leq g^{-1}(y)\right) = F_X\left(g^{-1}(y)\right) \end{aligned}$$

  - ▶ The density of Y is given by:

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) = \frac{\partial}{\partial y} F_X\left(g^{-1}(y)\right) \\ &= f_X\left(g^{-1}(y)\right) \cdot \frac{\partial}{\partial y} g^{-1}(y). \end{aligned}$$

- Suppose g() is differentiable and strictly **decreasing**:
  - ▶ The CDF of Y can be derived using:

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr(g(X) \leq y) \\ &= \Pr\left(X \geq g^{-1}(y)\right) = 1 - F_X\left(g^{-1}(y)\right) \end{aligned}$$

  - ▶ The density of Y is given by:

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) = \frac{\partial}{\partial y}\left(1 - F_X\left(g^{-1}(y)\right)\right) \\ &= -f_X\left(g^{-1}(y)\right) \cdot \frac{\partial}{\partial y} g^{-1}(y). \end{aligned}$$

- Essentially, if g() is a strictly monotonic function:

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \cdot \left|\frac{\partial}{\partial y} g^{-1}(y)\right|$$

- Common Transformations:

3. Exponential transformation: $Y = e^{a \cdot X}$, $a > 0$:

$$g^{-1}(y) = \frac{\log(y)}{a}, \qquad F_Y(y) = F_X\left(\frac{\log(y)}{a}\right),$$
$$\left|\frac{\partial}{\partial y} g^{-1}(y)\right| = \frac{1}{a \cdot y}, \qquad f_Y(y) = \frac{f_X\left(\frac{\log(y)}{a}\right)}{a \cdot y}.$$

4. Inverse transformation: $Y = 1/X$, $x > 0$:

$$g^{-1}(y) = \frac{1}{y}, \qquad F_Y(y) = 1 - F_X\left(\frac{1}{y}\right),$$
$$\left|\frac{\partial}{\partial y} g^{-1}(y)\right| = \frac{1}{y^2}, \qquad f_Y(y) = \frac{f_X\left(\frac{1}{y}\right)}{y^2}.$$

Example: Let $Y = -3X + 4$. Find $F_Y(y)$ and $f_Y(y)$ if $X \sim \text{U}(1, 9)$. $\cdot \left|y^{\frac{1}{n}-1}\right|$.

Solution: We know that:

$$f_X(x) = \begin{cases} 0, & \text{if } x < 1 \text{ or } x > 9; \\ 1/8, & \text{if } 1 \leq x \leq 9. \end{cases}$$
$$F_X(x) = \begin{cases} 0, & \text{if } x < 1; \\ \frac{x-1}{8}, & \text{if } 1 \leq x \leq 9; \\ 1, & \text{if } x > 9. \end{cases}$$

Support of Y: $g(1) = -3 \cdot 1 + 4 = 1$ and $g(9) = -3 \cdot 9 + 4 = -23$. We have $g^{-1}(Y) = -(Y-4)/3$.

Applying the CDF technique,

$$\begin{aligned} F_Y(y) =& 1 - F_X\left(g^{-1}(y)\right) = 1 - F_X\left(-\frac{y-4}{3}\right) \\ =& 1 - \frac{-\frac{y-4}{3} - 1}{8} = 1 + \frac{y-1}{24} \\ =& \frac{y+23}{24}, \quad \text{if } -23 \leq y \leq 1, \end{aligned}$$

and zero if $y < -23$ and one if $y > 1$. Differentiating $F_Y(y)$ gives

$$f_Y(y) = \frac{1}{24}, \qquad \text{if } -23 \leq y \leq 1,$$

and zero if $y \in (-\infty, -23) \cup (1, \infty)$.

The Jacobian Transformation Technique

- Consider two continuous RVs X_1 and X_2. We are given:
  $u_1 = g_1(x_1, x_2) \quad$ and $\quad u_2 = g_2(x_1, x_2)$
- We can find two functions:
  $x_1 = h_1(u_1, u_2) \quad$ and $\quad x_2 = h_2(u_1, u_2)$
    - Note: We are interested in the joint distribution of U_1 and U_2

- The Jacobian of this transformation is:

$$J(u_1, u_2) = \det\left(\begin{bmatrix} \frac{\partial h_1(u_1, u_2)}{\partial u_1} & \frac{\partial h_1(u_1, u_2)}{\partial u_2} \\ \frac{\partial h_2(u_1, u_2)}{\partial u_1} & \frac{\partial h_2(u_1, u_2)}{\partial u_2} \end{bmatrix}\right)$$
$$= \frac{\partial h_1(u_1, u_2)}{\partial u_1} \cdot \frac{\partial h_2(u_1, u_2)}{\partial u_2} - \frac{\partial h_2(u_1, u_2)}{\partial u_1} \cdot \frac{\partial h_1(u_1, u_2)}{\partial u_2},$$

- The joint density of U_1 and U_2 is given by:
  $f_{U_1,U_2}(u_1, u_2)$
  $= f_{X_1,X_2}(h_1(u_1, u_2), h_2(u_1, u_2)) \cdot |J(u_1, u_2)|.$
- Summarised Procedure:
  1. Start with $u_1 = g_1(x_1, x_2)$ and $u_2 = g_2(x_1, x_2)$;
  2. Find functions h_1, h_2 such that $h_1(g_1(x_1, x_2), g_2(x_1, x_2)) = x_1$ and $h_2(g_1(x_1, x_2), g_2(x_1, x_2)) = x_2$;
  3. Find the absolute value of the Jacobian, $|J(u_1, u_2)|$;
  4. Multiply that with the joint density of X_1, X_2 evaluated in $h_1(u_1, u_2)$, $h_2(u_1, u_2)$.

Example:

Let $\{X_1, X_2\}$ be uncertainty in the claim size of home insurance and unemployment insurance. We have the joint p.d.f.:

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} \exp(-(x_1 + x_2)), & \text{if } x_1 \geq 0 \text{ and } x_2 \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Question: Find the covariance between the aggregate claim size and the proportion due to home insurance.

Solution: Find the joint density between

$$Y_1 = X_1 + X_2 \quad \text{and} \quad Y_2 = \frac{X_1}{X_1 + X_2}.$$

1. We have transformations: $Y_1 = X_1 + X_2$ and $Y_2 = \frac{X_1}{X_1+X_2}$.
   Note that the domain for Y_1 is $[0, \infty)$ and Y_2 is $[0, 1]$.
2. Thus the inverses are:

$$X_1 = Y_2 \cdot (X_1 + X_2) = Y_1 \cdot Y_2$$
$$X_2 = Y_1 - X_1 = Y_1 \cdot (1 - Y_2)$$

3. The Jacobian of this transformation is

$$J(y_1, y_2) = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix} = |-y_2 y_1 - y_1(1 - y_2)| = y_1.$$

4. For $y_1 \geq 0$ and $0 \leq y_2 \leq 1$ we have:

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= \exp\left(-(y_1 \cdot y_2 + y_1 \cdot ((1-y_2))\right) \cdot y_1 \\ &= \exp\left(-y_1\right) \cdot y_1. \end{aligned}$$

Hence:

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \exp\left(-y_1\right) \cdot y_1, & \text{for } y_1 \geq 0 \text{ and } 0 \leq y_2 \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

Hence, covariance equals zero (independent!).

The MGF Technique

- Important Properties of the MGF:

$$\begin{aligned} M_{mX+b}(t) =& \mathbb{E}\left[e^{t\cdot(mX+b)}\right] \\ =& \mathbb{E}\left[\exp\left(t \cdot mX\right) \cdot \exp\left(bt\right)\right] \\ =& \mathbb{E}\left[\exp\left(tm \cdot X\right)\right] \cdot \exp\left(bt\right) \\ =& M_X\left(mt\right) \cdot \exp\left(bt\right) \end{aligned}$$

$$\begin{aligned} M_{X+Y}(t) =& \mathbb{E}\left[e^{t\cdot(X+Y)}\right] \\ =& \mathbb{E}\left[\exp\left(t \cdot X\right) \cdot \exp\left(t \cdot Y\right)\right] \\ =& \mathbb{E}\left[\exp\left(t \cdot X\right)\right] \cdot \mathbb{E}\left[\exp\left(t \cdot Y\right)\right] \\ =& M_X(t) \cdot M_Y(t) \end{aligned}$$

- This method is effective when we can obtain the MGF because it determines a unique distribution. Suppose we are interested in the distribution of:

$$U = g\left(X_1, \ldots, X_n\right)$$

- Determine the distribution of U by finding its MGF:

$$\begin{aligned} M_U\left(t\right) =& \mathbb{E}\left[e^{Ut}\right] \\ =& \int_{-\infty}^{\infty} \cdot \ldots \cdot \int_{-\infty}^{\infty} e^{g(x_1,\ldots,x_n)t} f_{X_1,\ldots,X_n}\left(x_1, \ldots, x_n\right) dx_1 \ldots dx_n. \end{aligned}$$

- Compare the MGF to that of a known distribution in order to determine the distribution of U.

- In the case where:

$$U = b_1 \cdot X_1 + \ldots + b_n \cdot X_n,$$

- We have:

$$\begin{aligned} M_U\left(t\right) =& \mathbb{E}\left[e^{(b_1\cdot X_1+\ldots+b_n\cdot X_n)\cdot t}\right] = \mathbb{E}\left[e^{X_1\cdot b_1\cdot t}\right] \cdot \ldots \cdot \mathbb{E}\left[e^{X_n\cdot b_n\cdot t}\right] \\ =& M_{X_1}\left(b_1 \cdot t\right) \cdot \ldots \cdot M_{X_n}\left(b_n \cdot t\right). \end{aligned}$$

The m.g.f. of U is the product of the m.g.f. of $X_1, \ldots, X_n$.

Example:

Let X_i be the i.i.d. claims arriving from males motor vehicle

Solution: Let $X_i \sim \text{Poisson}(\lambda_1)$ and $Y_i \sim \text{Poisson}(\lambda_2)$, where X_1, X_2 are independent. Let $U = \sum_{i=1}^{n} X_i + \sum_{i=1}^{m} Y_i$ denote the total number of claims. The m.g.f. of U is given by:

$$M_U(t) = \prod_{i=1}^{n} M_{X_i}(t) \cdot \prod_{i=1}^{m} M_{Y_i}(t)$$
$$= \left(e^{\lambda_1 \cdot (e^t - 1)}\right)^n \cdot \left(e^{\lambda_2 \cdot (e^t - 1)}\right)^m = e^{(n \cdot \lambda_1 + m \cdot \lambda_2) \cdot (e^t - 1)},$$

which is the m.g.f. of a Poisson with parameter $n \cdot \lambda_1 + m \cdot \lambda_2$.

$$U \sim \text{Poisson}(n \cdot \lambda_1 + m \cdot \lambda_2).$$

Question: Find the distribution of the total number of claims.

Convolutions

- Suppose *X* and *Y* are discrete and independent. Let *Z* = *X* + *Y*, then:

$$p_Z(z) = \sum_{x=0}^{z} p_X(x) \cdot p_Y(z - x)$$

  - This is called the convolution of p_x *and* p_y
- Suppose *X* and *Y* are continuous and independent. Let *Z* = *X* + *Y*, then:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x)\, dx.$$

Example:

Let $X_i \sim \text{EXP}(\lambda)$ be the size of the semiannual expected discounted value of newly issued long-term disability insurance claims $(i = 1, 2)$:

$$f_{X_i}(x_i) = \begin{cases} \lambda \cdot \exp(-\lambda \cdot x_i), & \text{if } x_i \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Question: Find the distribution of the annual claim size.

Solution: Let $Z = X_1 + X_2$. If $z \geq 0$ we have:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X_1}(z - x_2) \cdot f_{X_2}(x_2) dx_2$$
$$= \int_0^z \lambda \cdot \exp(-\lambda \cdot (z - x_2)) \cdot \lambda \cdot \exp(-\lambda \cdot x_2)\, dx_2$$
$$= \int_0^z \lambda^2 \cdot \exp(-\lambda \cdot z)\, dx_2$$
$$= \lambda^2 \cdot z \cdot \exp(-\lambda \cdot z).$$

The Chi-Squared Distribution

- Suppose $Z \sim N(0,1)$, then:

$$Y = Z^2 \sim \chi^2(1)$$

  - has a chi-squared distribution with one degree of freedom

- PDF:

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} \cdot e^{-\frac{y}{2}}, \quad y \geq 0.$$

- CDF:

$$\begin{aligned} F_Y(y) =& P(Z^2 \leq y) \\ =& P(-\sqrt{y} \leq Z \leq \sqrt{y}) \\ =& F_Z(\sqrt{y}) - F_Z(-\sqrt{y}) = 2 \cdot F_Z(\sqrt{y}) - 1 \end{aligned}$$

- Expected Value:

$$\mathbb{E}[Y] = \mathbb{E}\left[Z^2\right] = 1.$$

- Variance:

$$Var(Y) = \mathbb{E}\left[Y^2\right] - (\mathbb{E}[Y])^2 = \mathbb{E}\left[Z^4\right] - \left(\mathbb{E}\left[Z^2\right]\right)^2 = 3 - 1 = 2.$$

- Now consider:

$$X = \sum_{i=1}^{n} Z_i^2.$$

- X has a chi-squared distribution with n degrees of freedom.
- PDF:

$$f_X(x) = \frac{1}{2^{n/2} \cdot \Gamma(n/2)} \cdot x^{(n-2)/2} \cdot e^{-x/2}, \quad \text{if } x > 0,$$

  - Recall the p.d.f. of a Chi-squared with one degree of freedom Y:

$$f_Y(y) = \frac{1}{\sqrt{2} \cdot \Gamma(1/2)} \cdot y^{-1/2} \cdot e^{y/2},$$

    which has a Gamma$\left(\frac{1}{2}, \frac{1}{2}\right)$.
  - For independent $Y_1, Y_2, \ldots, Y_n \sim \chi^2(1)$,

$$Y_1 + Y_2 + \ldots + Y_n \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right) \stackrel{\text{dist}}{=} \chi^2(n).$$

  - The sum of i.i.d. Gamma random variables Gamma(α_i, λ) is also a Gamma random variable but with Gamma$(\sum_{i=1}^{n} \alpha_i, \lambda)$.
- Expected Value:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} Y_i\right] \stackrel{*}{=} n \cdot \mathbb{E}[Y_i] = n.$$

- Variance:

$$Var(X) = Var\left(\sum_{i=1}^{n} Y_i\right) \stackrel{*}{=} n \cdot Var(Y_i) = 2 \cdot n.$$

- MGF:

$$M_X(t) = (1 - 2 \cdot t)^{-n/2}, \quad t < 1/2,$$

Student-t Distribution

- Suppose $Z \sim N(0,1)$ and $V \sim \chi^2(r) = \sum_{k=1}^{r} Z_i^2$, Z_i i.i.d. and Z, V are independent. Then, the random variable:

$$T = \frac{Z}{\sqrt{V/r}}$$

has a t-distribution with r degrees of freedom.

The F Distribution

- Suppose $U \sim \chi^2(n_1)$ and $V \sim \chi^2(n_2)$ are two independent chi squared distributed random variables.

- Then, the random variable:

$$F = \frac{U/n_1}{V/n_2}$$

has a F distribution with n_1 and n_2 degrees of freedom.

Properties of Sample Mean and Sample Variance

- Sample mean:

$$\overline{X} = \frac{1}{n} \cdot \sum_{k=1}^{n} X_k,$$

- Sample Variance:

$$S^2 = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (X_k - \overline{X})^2.$$

- If the sample is Normal, then the following properties hold:
  - $\overline{X} \sim \mathrm{N}\left(\mu, \frac{1}{n}\sigma^2\right)$
  - $T = \dfrac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$
  - $\dfrac{(n-1) \cdot S^2}{\sigma^2} \sim \chi^2(n-1)$
  - $\overline{X}$ and S^2 are independent

Estimation Techniques

Definition of an Estimator

- When a density has an unknown parameter θ, we estimate the parameter based on a random sample.
- Any statistic whose values are used to estimate θ is called an **estimator** of θ, often denoted $\hat{\theta}$.
- A specific value of the estimator based on observed sample values is called an **estimate**.
    - For example:

$$\begin{aligned} T(X_1, X_2, \ldots, X_n) = \overline{X} &= \frac{1}{n}\sum_{j=1}^{n} X_j, && \text{estimator;} \\ \widehat{\theta} &= 0.23, && \text{point estimate.} \end{aligned}$$

The Method of Moments

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with density $f_X(\cdot|\underline{\theta})$ and a set of k parameters $\underline{\theta} = [\theta_1, \theta_2, \ldots, \theta_k]^\top$.

- The Method of Moments procedure is:
    1. Equate (the first) k sample moments to the corresponding k population moments;
    2. Express the k population moments as functions of the parameters $\theta_1, \theta_2, \ldots, \theta_k$ of the distribution f_X;
    3. Solve the resulting system of simultaneous equations in order to express the parameters as function of the moments

- Denote the **sample moments** by:

$$m_1 = \frac{1}{n} \cdot \sum_{j=1}^{n} x_j, \quad m_2 = \frac{1}{n} \cdot \sum_{j=1}^{n} x_j^2, \quad \ldots, \quad m_k = \frac{1}{n} \cdot \sum_{j=1}^{n} x_j^k,$$

- and the **population moments** by:

$$\mu_1(\theta_1, \theta_2, \ldots, \theta_k) = \mathbb{E}[X], \quad \mu_2(\theta_1, \theta_2, \ldots, \theta_k) = \mathbb{E}\left[X^2\right], \\ \ldots \quad , \mu_k(\theta_1, \theta_2, \ldots, \theta_k) = \mathbb{E}\left[X^k\right].$$

- The system of equations to solve for $(\theta_1, \theta_2, \ldots, \theta_k)$ is given by:

$$m_j = \mu_j(\theta_1, \theta_2, \ldots, \theta_k), \quad \text{for } j = 1, 2, \ldots, k.$$

Solving this provides us the point estimate $\widehat{\underline{\theta}}$.

Example:

Maximum Likelihood Estimation

- The maximum likelihood estimator is another estimator.
- The likelihood function is given by:
  - Let $L(\theta) = L(\theta; x_1, x_2, \ldots, x_n)$ be the likelihood function for $X_1, X_2, \ldots, X_n$.
  - The set of parameters $\widehat{\theta} = \widehat{\theta}(x_1, x_2, \ldots, x_n)$ (note: function of observed values) that maximizes $L(\theta)$ is the maximum likelihood estimate of θ.
  - The random variable $\widehat{\theta}(X_1, X_2, \ldots, X_n)$ is called the maximum likelihood estimator.
- When $X_1, X_2, \ldots, X_n$ is a random sample from $f_x(x|\theta)$, then the likelihood function is:

$$L(\theta; x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} f_X(x_j|\theta)$$

- We can optimise the likelihood function, but it is easier to optimise the log-likelihood function (they achieve the same result).
- We define the log-likelihood function as:

$$\begin{aligned}\ell(\theta_1, \theta_2, \ldots, \theta_k; \underline{x}) &= \log(L(\theta_1, \theta_2, \ldots, \theta_k; \underline{x})) \\ &= \log\left(\prod_{j=1}^{n} f_X(x_j|\theta)\right) \\ &\overset{*}{=} \sum_{j=1}^{n} \log(f_X(x_j|\theta)).\end{aligned}$$

Suppos)
distribu

Questi
estimat

Solution: Equate population moment to sample moment:

1. Determine the likelihood function $L(\theta_1, \theta_2, \ldots, \theta_k; \underline{x})$;
2. Determine the log-likelihood function $\ell(\theta_1, \theta_2, \ldots, \theta_k; \underline{x}) = \log(L(\theta_1, \theta_2, \ldots, \theta_k; \underline{x}))$;
3. The First Order Condition (FOC): Equate the derivatives of $\ell(\theta_1, \theta_2, \ldots, \theta_k; \underline{x})$ w.r.t. $\theta_1, \theta_2, \ldots, \theta_k$ to zero ($\Rightarrow$ critical point: global/local minimum/maximum/saddle point).
4. The Second Order Condition (SOC): Check whether the Hessian at the critical point is negative (one variable) or negative-definite (many variables) : if yes you have a maximum

Equate
represe

Then tl

- The general procedure to find the ML estimator is:

Example:

Now we need to maximize this log-likelihood function with respect to the parameter λ.

3. Taking the first order condition (FOC) with respect to λ we have:

$$\frac{\partial}{\partial \lambda}\ell(\lambda) = 0 \qquad \Rightarrow \qquad -n + \frac{1}{\lambda}\sum_{k=1}^{n} x_k = 0.$$

This gives the maximum likelihood estimate (MLE):

$$\widehat{\lambda} = \frac{1}{n}\sum_{k=1}^{n} x_k = \overline{x},$$

which equals to the sample mean.

4. Check for second derivative condition to ensure global maximum.

Bayesian Estimator

- In the Bayesian approach, we consider the parameter Θ to be a random variable itself (rather than deterministic), with a **prior density** denoted $\pi(\theta)$
- Bayesian Interpretation: You have a prior belief about the distribution of the variable Θ. The you observe some data $X_1, X_2, \ldots, X_n$ which is related to Θ and the update your belief.
  - This new estimated distribution is called the **posterior distribution**.
- ▶ A sample $\underline{X} = \underline{x} (= [x_1, x_2, \ldots, x_T]^{\top})$ is taken from its population and the prior density is updated using the information drawn from this sample and applying Bayes' rule. This updated prior is called the posterior density, which is the conditional density of Θ given the sample $\underline{X} = \underline{x}$, i.e. $\pi(\theta|\underline{x})$ $(\equiv f_{\Theta|\underline{X}}(\theta|\underline{x}))$.
- ▶ So we're using a conditional r.v., $\Theta|\underline{X}$, associated with the multivariate distribution of Θ and the $\underline{X}$.
- ▶ We have that the posterior is given by:

$$\pi(\theta|\underline{x}) \propto f_{\underline{X}|\Theta}(x_1, x_2, \ldots, x_T|\theta) \cdot \pi(\theta). \qquad (2)$$

Either use equation (1) (difficult/tidious integral!) or (2).

Equation (2) can be used to find the posterior density by:

I. Find c such that $c \cdot \int f_{\underline{X}|\Theta}(x_1, x_2, \ldots, x_T|\theta) \cdot \pi(\theta)\, d\theta = 1$.

II. Find a (special) distribution that is proportional to $f_{\underline{X}|\Theta}(x_1, x_2, \ldots, x_T|\theta) \cdot \pi(\theta)$. (fastest way, if possible!)

- Summarised Procedure:
  1. Start by writing down $\pi(\theta)$, $f_{\underline{X}}(\underline{x})$ and $f_{\underline{X}|\Theta}(\underline{x}|\theta)$.
  2. Find the posterior density $\pi(\theta|\underline{x})$ using (1) (difficult/tedious integral!) or (2). This is your updated distribution for Θ.
  3. (Optional) If you want an estimator of Θ, you can compute the Bayesian estimator $\mathbb{E}[\Theta|\underline{X}]$: this is the mean of Θ, under the updated distribution [2].

Example:

Let $X_1, X_2, \ldots, X_T$ be i.i.d. Bernoulli(Θ), i.e., $(X_i|\Theta = \theta) \sim \text{Bernoulli}(\theta)$. Assume the **prior density** of Θ is Beta(a, b) so that:

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\cdot\Gamma(b)} \cdot \theta^{a-1} \cdot (1-\theta)^{b-1}.$$

We know that the conditional density (density conditional on knowing the value of θ) of our data is given by:

$$\begin{aligned} f_{\underline{X}|\Theta}(\underline{x}|\theta) =& \theta^{x_1}(1-\theta)^{1-x_1} \cdot \theta^{x_2}(1-\theta)^{1-x_2} \cdot \ldots \cdot \theta^{x_T}(1-\theta)^{1-x_T} \\ =& \theta^{\sum_{j=1}^{T} x_j} \cdot (1-\theta)^{T-\sum_{j=1}^{T} x_j} \stackrel{*}{=} \theta^s \cdot (1-\theta)^{T-s}. \end{aligned}$$

This is just the likelihood function.
* Simplifying notation, let $s = \sum_{j=1}^{T} x_j$.

Easy method: The **posterior density**, the density of Θ given $\underline{X} = \underline{x}$ is proportional to:

$$\begin{aligned} \pi(\theta|\underline{x}) \propto& f_{\underline{X}|\Theta}(x_1, x_2, \ldots, x_T|\theta) \cdot \pi(\theta) \\ =& \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \theta^{(a+s)-1} \cdot (1-\theta)^{(b+T-s)-1} \end{aligned} \qquad (3)$$

I. **Posterior density** is also solvable by finding c such that:

$$\int c \cdot \frac{\Gamma(a+b)}{\Gamma(a)\cdot\Gamma(b)} \cdot \theta^{(a+s)-1} \cdot (1-\theta)^{(b+T-s)-1}\, d\theta = 1.$$

Posterior density is $c \cdot f_{\underline{X}|\Theta}(x_1, x_2, \ldots, x_T|\theta) \cdot \pi(\theta)$.

II. However, we observe (3) is proportional to the p.d.f. of a Beta$(a+s, b+T-s)$: this is our posterior!

To get a point estimate, we take the mean of this r.v. with the above posterior density:

$$\widehat{\theta}^B = \mathbb{E}[\Theta|\underline{X} = \underline{x}] = \mathbb{E}[\Xi \sim \text{Beta}(a+s, b+T-s)] = \frac{a+s}{a+b+T}$$

which gives the **Bayesian estimator** of Θ.

Limit Theorems

Chebyshev's Inequality

- Chebyshev's Inequality (works for ANY distribution):

$$\Pr(|X-\mu| > \epsilon) \le \frac{\sigma^2}{\epsilon^2}.$$

Interesting example: set $\epsilon = k \cdot \sigma$ then:

$$\Pr(|X-\mu| > k\cdot\sigma) \le \frac{1}{k^2}.$$

Example:

The distribution of fire insurance claims does not have a special distribution. We do know that the mean claim size in the portfolio is $50 million with a standard deviation of $150 million.

Question: What is the upper bound for the probability that the claim size is larger than $500 million.

Solution: With $k = 3$, we have:

$$\begin{aligned}\Pr(X > 500) &= \Pr(X - 50 > 3 \cdot 150) \\ &= \Pr(X - \mu > k \cdot \sigma) \\ &\leq \Pr(|X - \mu| > k \cdot \sigma) \\ &\leq \frac{1}{k^2} = \frac{1}{9}.\end{aligned}$$

Thus, $\Pr(X > 500) \leq 1/9$.

Convergence Concepts

- Suppose $X_1, X_2, \ldots$ form a sequence of random variables.
- We say X_n **converges in probability** to the random variable X as $n \to \infty$ if and only if:

  $$\Pr(|X_n - X| > \epsilon) \to 0, \qquad \text{as } n \to \infty,$$

  - We write:

    $$X_n \xrightarrow{p} X, \text{ as } n \to \infty.$$
- We say X_n **converges in density** to the random variable X as $n \to \infty$ if and only if:

  $$F_{X_n}(x) \to F_X(x), \qquad \text{as } n \to \infty$$

  - We write:

    $$X_n \xrightarrow{d} X, \text{ as } n \to \infty$$
  - This is usually referred to as *weak convergence*

Law of Large Numbers

- Let $X_1, X_2, \ldots, X_n$ be independent random variables with common mean $\mathbb{E}[X_k] = \mu$, common variance $Var(X_k) = \sigma^2$, for $k = 1, 2, \ldots, n$.
- Define the sequence of sample means as:

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

- The (weak) law of large numbers, states that

$\overline{X}_n$ converges in probability to μ.

- Stated otherwise, the (weak) law of large numbers (LLN) says that:

$$\Pr\left(\left|\overline{X}_n - \mu\right| > \epsilon\right) \to 0, \qquad \text{as } n \to \infty.$$

- Interpretation: as n gets bigger and bigger, the sample mean gets closer and closer to the theoretical mean μ.

Central Limit Theorem

- Suppose $X_1, X_2, \ldots, X_n$ are independent, identically distributed random variables with finite mean μ and finite variance σ^2. As before, denote the sample mean by $\overline{X}_n$.

- Then, the central limit theorem states:

$$\frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} N(0, 1), \qquad \text{as } n \to \infty.$$

This holds for <u>all</u> r.v. with finite mean and variance, not only normal r.v.!

- We can write this result as:

$$\lim_{n \to \infty} \Pr\left(\frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \leq x\right) = \Phi(x),$$

for all x where $\Phi(\cdot)$ denotes the cdf of a standard normal r.v..

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

is approximately standard normally distributed if n is 'large'

- The Central Limit Theorem is usually expressed in terms of the standardized sums $S_n = \sum_{k=1}^{n} X_k$. Then the CLT applies to the random variable:

$$Z_n = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} \xrightarrow{d} \mathrm{N}(0, 1), \qquad \text{as } n \to \infty.$$

Example:

An insurer offers builder's risk insurance. It has nearly 400 contracts and has been offering the same product for many years. From past experience, the sample mean of a claim is $10 million and the sample standard deviation is $25 million.

Question: What is the probability that in a year the claim size is larger than \$5 billion?

Solution: Using CLT (why is $\sigma \approx$ sample s.d.?)

$$\frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} N(0,1), \quad \text{as } n \to \infty$$

$$\Rightarrow \overline{X}_n \sim \mathrm{N}\left(\mu, (\sigma/\sqrt{n})^2\right) \Rightarrow n \cdot \overline{X}_n \sim \mathrm{N}\left(n \cdot \mu, n \cdot \sigma^2\right)$$

Let X be the total loss, then we have

$$X = 400 \cdot \overline{X}_{400} \sim N(400 \cdot 10, 400 \cdot 25^2).$$

The probability we need to calculate is

$$P(X > 5000) = P(Z > \frac{5000 - 4000}{20 \cdot 25}) = 1 - 0.9772 = 0.0228$$

Normal Approximation to the Binomial

- We know: a Binomial random variable is the sum of Bernoulli random variables. Let $X_k \sim$ Bernoulli (p). Then:

  $$S = X_1 + X_2 + \ldots + X_n$$

  has a Binomial(n, p) distribution.

- Applying the Central Limit Theorem, S must be approximately normal with mean $\mathbb{E}[S] = n \cdot p$ and variance $Var(S) = n \cdot p \cdot q$ so that, approximately, for large n we have:

  $$\frac{S - n \cdot p}{\sqrt{n \cdot p \cdot q}} \sim \mathrm{N}(0,1).$$

- Continuity correction for binomial: note that Binomial random variable X takes integer values $k = 0, 1, 2, \ldots$ but Normal probability is continuous so that for value:

$$\Pr\left(X = k\right),$$

we require the Normal approximation:

$$\Pr\left(\frac{\left(k - \frac{1}{2}\right) - \mu}{\sigma} < Z < \frac{\left(k + \frac{1}{2}\right) - \mu}{\sigma}\right)$$

Normal Approximation to the Poisson

- Approximation of Poisson by Normal for large values of λ.

  Let X_n be a sequence of Poisson random variables with increasing parameters $\lambda_1, \lambda_2, \ldots$ such that $\lambda_n \to \infty$.

  We have:

$$\begin{aligned} \mathbb{E}[X_n] =& \lambda_n \\ Var(X_n) =& \lambda_n \end{aligned}$$

- Standardize the random variable (i.e., subtract mean and divide by standard deviation):

$$Z_n = \frac{X_n - \mathbb{E}[X_n]}{\sqrt{Var(X_n)}} = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}} \xrightarrow{d} Z \sim \mathrm{N}(0, 1).$$

Evaluating Estimators

Methods to Evaluate Estimators

- Estimators can be compared using the following criteria:
  - Bias
  - Variance
  - Mean squared error
  - Consistency
- Among unbiased estimators, it is sometimes possible to find one with the lowest possible variance, called the **Minimum Variance Unbiased Estimator** (MVUE)

- Mean Squared Error and Bias:

  The mean squared error (MSE) of an estimator

  Consider two unbiased estimators, say T_1 and T_2. We define efficiency of T_1 relative to T_2 as:

$$eff\,(T_1, T_2) = \frac{Var(T_2)}{Var(T_1)}.$$

  It is clear that if this is larger than 1, then:

$$\mathbb{E}[T]^2$$

$$Var(T_2) > Var(T_1),$$

  i.e., estimator T_1 has lower variance than estimator T_2.

  ,

  Thus a high value of $eff(T_1, T_2)$ implies we prefer T_1 to T_2.

  where $\text{Bias}(T) = \mathbb{E}[T] - \theta$, note θ is a constant.

  An unbiased estimator has: $\mathbb{E}[T] = \theta$.

- Efficiency:
- An estimator *T* is said to be a best unbiased estimator (MVUE) of θ if it satisfies two conditions:

Cramer-Rao Lower Bound (CRLB)

- A lower bound of the variance for unbiased estimators is:

$$Var(T(X_1, X_2, \dots, X_n)) \geq \frac{1}{n \cdot I_{f^\star}(\theta)},$$

- The Fisher information of the parameter θ is defined to be the function:

$$\begin{aligned} I_{f^\star}(\theta) &= \mathbb{E}\left[\left(\frac{\partial \log(f_X(X|\theta))}{\partial \theta}\right)^2\right] \overset{*}{=} -\mathbb{E}\left[\frac{\partial^2 \log(f_X(X|\theta))}{\partial \theta^2}\right] \\ &\overset{**}{=} \mathbb{E}\left[\left(\frac{\partial \ell(\underline{X};\theta)}{\partial \theta}\right)^2\right] / n^2 \overset{**}{=} -\mathbb{E}\left[\frac{\partial^2 \ell(\underline{X};\theta)}{\partial \theta^2}\right] / n, \end{aligned}$$

  - The estimator T has the smallest variance, i.e.,

Given n draws from a $\text{Bin}(m, p)$ r.v. with $f_X(x;p) = \binom{m}{x} \cdot p^x \cdot (1-p)^{m-x}$

$$\log\left(f_X(x;p)\right) = \log\left(\binom{m}{x}\right) + x \cdot \log(p) + (m - x) \cdot \log(1-p)$$

Question: Find the CRLB for an estimator of p (if m is known)

Solution: First, Fisher information (* $Var(X) = mp(1-p)$):

$$\frac{\partial^2 \log\left(f_X(x;p)\right)}{\partial p^2} = \frac{-x}{p^2} - \frac{m-x}{(1-p)^2}$$

$$I_{f^\star}(p) = -\mathbb{E}\left[\frac{\partial^2 \log\left(f_X(X;p)\right)}{\partial p^2}\right] = -\left(\frac{-\mathbb{E}[X]}{p^2} - \frac{m - \mathbb{E}[X]}{(1-p)^2}\right) = \frac{m}{p(1-p)}.$$

Thus, the Cramér-Rao Lower Bound is given by:

$$Var(T(X_1, \dots, X_n)) \geq \frac{1}{n \cdot \frac{m}{p(1-p)}} = \frac{p(1-p)}{m \cdot n}.$$

Example:

Consistency

- A sequence of estimators $\{T_n\}$ is a consistent sequence of estimators of the parameter θ if:
  $\lim_{n\to\infty} \Pr(|T_n - \theta| < \epsilon) = 1,$
  i.e., $T_n \xrightarrow{p} \theta.$
- Equivalently, if T_n is a sequence of estimators that satisfies the following conditions, then it is a sequence of consistent estimators:
  i) $\lim_{n\to\infty} Var(T_n) = 0$ (the uncertainty in the estimate is zero as $n \to \infty$);
  ii) $\lim_{n\to\infty} Bias(T_n) = 0$ (estimator is asymptotically unbiased);

Properties of MLE

Then the $\widehat{\theta}_n$ are asymptotically normally distributed with mean:

$$\mathbb{E}\left[\widehat{\theta}_n\right] \to \theta, \qquad \text{as } n \to \infty$$

and variance:

$$Var\left(\sqrt{n}\cdot\widehat{\theta}_n\right) \to \frac{1}{I_{f^\star}(\theta)}, \qquad \text{as } n \to \infty.$$

For large n, we then have an approximate variance of the MLE $Var\left(\widehat{\theta}_n\right) \approx \left(n \cdot I_{f^\star}(\theta)\right)^{-1}$
We write this as:

$$\sqrt{n}\cdot\left(\widehat{\theta}_n - \theta\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{1}{I_{f^\star}(\theta)}\right).$$

For functions of the parameter, say $g(\theta)$, we can easily extend the theorem, except there is a delta-method adjustment to the variance.

Thus, assuming $g(\theta)$ is a differentiable function of θ, then:

$$\sqrt{n}\left(g\left(\widehat{\theta}_n\right) - g(\theta)\right) \xrightarrow{d} N\left(0, \left(g'(\theta)\right)^2 \cdot \frac{1}{I_{f^\star}(\theta)}\right),$$

where $g'(\theta)$ is the first derivative of g with respect to the parameter θ.

CI For Maximum Likelihood Estimates

$$\sqrt{n\cdot I_{f^\star}(\theta)}\cdot\left(\widehat{\theta} - \theta\right) \xrightarrow{d} \mathrm{N}(0,1).$$

Using this as a pivotal quantity, we have approximately:

$$\Pr\left(-z_{1-\alpha/2} < \sqrt{n\cdot I_{f^\star}\left(\widehat{\theta}\right)}\cdot\left(\widehat{\theta} - \theta\right) < z_{1-\alpha/2}\right) \approx 1 - \alpha,$$

or, equivalently:

$$\widehat{\theta} \pm z_{1-\alpha/2}\cdot\frac{1}{\sqrt{n\cdot I_{f^\star}\left(\widehat{\theta}\right)}}$$

is an approximate $100(1-\alpha)\%$ confidence interval for the parameter θ.

Building Hypothesis Tests

Hypothesis Testing

- Formal means of distinguishing between probability distributions using sample data.
- Hypothesis testing procedure:
  i. Define a statistical hypothesis.
     Note that this includes a confidence level (α);
  ii. Define the test statistic T (using past weeks knowledge);
  iii. Determine the rejection region $C^\star$;
  iv. Calculate the value of the test statistic, given observed data $(x_1, \ldots, x_n)$;
  v. Accept or reject H_0.
     Note: we assume that H_0 is true when testing! (see Type I and Type II errors)
- Null Hypothesis H_0: The hypothesis being tested
- Alternative Hypothesis H_1: the hypothesis accepted if H_0 is rejected.

Results of Decision Making		
	Hypothesis	
Decision	True	False
Reject H_0	Type I Error=α	No Error
Do not reject H_0	No Error	Type II Error=β

- Selection of the Null:
  - **Economic Method**: The null hypothesis is chosen so the consequences of incorrectly rejecting the null are the largest.
  - **Statistical Method**: Select the null as something you with to conclusively disprove.
- The level of significance α of the test can be determined from:

$$\Pr(T \in C^\star \,|\, H_0) = \alpha,$$

  - Where *T* is a test statistic and *C** is the rejection region.
  - It is the probability that you reject H_0 given that it is true
- Simple and Composite Hypotheses:
  - Simple
    $H_0 : \theta = \theta_0$ v.s. $H_1 : \theta = \theta_1$, for given constants θ_0 and θ_1.
  - Composite:
    $H_0 : \theta \in \omega_0$ v.s. $H_1 : \theta \in \omega_1$, for given sets ω_0 and ω_1.

Type 1 and Type 2 Errors

- Type 1: Null is true but rejected
- Type 2: Null is false but not rejected
- Probability of a Type 1 error: α

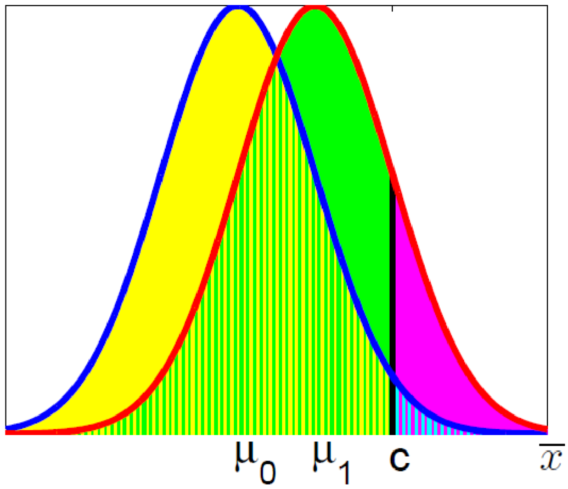
$$\alpha = \Pr(\text{Reject } H_0 \,|\, H_0 \text{ is true})$$

- Probability of a Type 2 error: β
- Power of a test: Probability that H_0 is rejected when it is false (= 1 – β)

$\beta = \Pr(\text{Do not reject } H_0 \,|\, H_0 \text{ is false})$

$$\pi(\theta) = \Pr((X_1, \ldots, X_n) \in C | \theta \in H_1).$$

  - Power function:
  - The value of the power function at a specific parameter point is called the *power of the test at that point*.



α (Type I error) is the blue shaded area
β (Type II error) is the green shaded area
$\pi(\mu_1)$ (power) is the purple shaded area

Example:
Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $N(\theta, 1)$ and that we wish to test

$$H_0 : \theta = 0 \quad \text{versus} \quad H_1 : \theta \neq 0.$$

Consider the rejection (or critical) region as:

$$C = \{(x_1, x_2, \ldots, x_n) : |\overline{x}| > 2\}$$

Question: Based on this information, calculate power, type I and II errors of this test.

We have that the level of significance is:

$$\begin{aligned}
\alpha &= \Pr((X_1, X_2, \ldots, X_n) \in C \,|H_0 \text{ is true}) \\
&= \Pr(|\overline{X}| > 2 \,|\theta = 0) \\
&= \Pr(\overline{X} > 2 \,|\theta = 0) + \Pr(\overline{X} < -2 \,|\theta = 0) \\
&= \Pr(\sqrt{n} \cdot \overline{X} > 2\sqrt{n}) + \Pr(\sqrt{n} \cdot \overline{X} < -2\sqrt{n}) \\
&\overset{*}{=} \Pr(Z > 2\sqrt{n}) + \Pr(Z < -2\sqrt{n}) \\
&= 2 \cdot \Phi(-2\sqrt{n}).
\end{aligned}$$

* using $\overline{X}|H_0 = \sum_{i=1}^{n} X_i/n \sim N(0, \frac{1}{n})$.
The power function is equal to the probability that the sample will fall within the critical region:

$$\begin{aligned}
\pi(\theta) &= \Pr((X_1, X_2, \ldots, X_n) \in C | \theta \in H_1) \\
&= \Pr(|\overline{X}| > 2 | \theta \in H_1) \\
&= \Pr(\overline{X} > 2 | \theta \in H_1) + \Pr(\overline{X} < -2 | \theta \in H_1) \\
&\overset{*}{=} \Pr(Z > \sqrt{n}(2 - \theta)) + \Pr(Z < \sqrt{n}(-2 - \theta)) \\
&= \Phi(-\sqrt{n}(2 - \theta)) + \Phi(\sqrt{n}(-2 - \theta)).
\end{aligned}$$

* using $\overline{X} = \sum_{i=1}^{n} X_i/n \sim N(\theta, \frac{\sigma^2}{n})$ (using m.g.f. technique).

p-value

- P-value is the probability that a value of the test statistic is as extreme as the observed value, given the null hypothesis is true

Example:

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $N(\theta, 1)$ and that we wish to test $H_0 : \theta = 0$ versus $H_a : \theta \neq 0$ with constant α as level of significance.

- Suppose the observed sample mean $\overline{X}$ turned out to be 3, we can calculate the p-value:

$$\begin{aligned} p\text{-value} &= \Pr\left(|\overline{X}| \geq 3 \,|\theta = 0\right) \\ &= \Pr\left(\overline{X} \geq 3 \,|\theta = 0\right) + \Pr\left(\overline{X} \leq -3 \,|\theta = 0\right) \\ &= \Pr\left(Z \geq 3\sqrt{n}\right) + \Pr\left(Z \leq -3\sqrt{n}\right) \end{aligned}$$

- If p-value is smaller then the level of significance $\implies$ reject the null hypothesis H_0.

Neyman-Pearson Lemma

- If $L(\underline{x}, \theta)$ denotes a likelihood of a sample, then the most powerful α level test is given by the rejection region:

$$\frac{L(\underline{x}, \theta_0)}{L(\underline{x}, \theta_1)} < k,$$

- Note: This method only works when the null and alternative are **simple**
- The likelihood ratio is denoted:

$$\Lambda(\underline{x}, \theta_0, \theta_1) = \frac{L(\underline{x}, \theta_0)}{L(\underline{x}, \theta_1)}$$

- The NPL gives the form of the rejection region, whereas the constant *k* depends on the value of α.

Example:

Let $X_1, X_2, \ldots, X_n$ from a normal distribution with known variance σ^2. Null and alternative hypotheses:

$$H_0 : \mu = \mu_0 \qquad \text{v.s.} \qquad H_1 : \mu = \mu_1$$

with α a (given) constant.

Question: What is the best critical region?

Solution:

The likelihood ratio is:

$$\begin{aligned}\Lambda(\underline{x};\theta_0,\theta_1) &= \frac{f_0(\underline{x})}{f_1(\underline{x})} \\ &= \frac{L(\underline{x};\theta_0)}{L(\underline{x};\theta_1)} = \frac{\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-1}{2\sigma^2}(x_i-\mu_0)^2\right)}{\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-1}{2\sigma^2}(x_i-\mu_1)^2\right)} \\ &= \frac{\exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\right)}{\exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_1)^2\right)} < k.\end{aligned}$$

Then:

Thus, by using a log-transformation (monotonic increasing transformation) we have that we reject H_0 when:

$$\begin{aligned}&\sum_{i=1}^{n}(x_i-\mu_1)^2-\sum_{i=1}^{n}(x_i-\mu_0)^2 < k_1 \qquad (=\log(k)\cdot 2\cdot\sigma^2) \\ \Rightarrow\ &\sum_{i=1}^{n}(x_i^2-2x_i\mu_1+\mu_1^2)-(x_i^2-2x_i\mu_0+\mu_0^2) < k_1 \\ \overset{*}{\Rightarrow}\ &2n\cdot\overline{x}(\mu_0-\mu_1)+n(\mu_1^2-\mu_0^2) < k_1 \\ \Rightarrow\ &\sum_{i=1}^{n}x_i\cdot(\mu_0-\mu_1) < k_2 \qquad (=(k_1-n(\mu_1^2-\mu_0^2))/2)\end{aligned}$$

From previous slide, rejection region is

$$\sum_{i=1}^{n}x_i\cdot(\mu_0-\mu_1) < k_2.$$

Therefore, we have two possibilities

(a) if $\mu_0>\mu_1\Rightarrow(\mu_0-\mu_1)>0$, the likelihood ratio is small if $\sum_{i=1}^{n}x_i$ (or $\overline{x}$) is small;

(b) if $\mu_0<\mu_1\Rightarrow(\mu_0-\mu_1)<0$, the likelihood ratio is small if $\sum_{i=1}^{n}x_i$ (or $\overline{x}$) is large.

In the latter (former) case need to determine a value x_0 such that, assuming that H_0 is true,

$$\begin{aligned}&\Pr\left(\overline{X}>x_0|H_0\right)=\alpha \\ &\Pr\left(\overline{X}<x_0|H_0\right)=\alpha\end{aligned}$$

We know: $\overline{X}|H_0\sim \mathrm{N}\left(\mu_0,\sigma^2/n\right)\Rightarrow$ calculate the value of x_0.

$$\begin{aligned}&\Pr\left(\overline{X}<x_0|H_0\right)=\alpha \quad\Rightarrow\quad \Pr\left(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}<\frac{x_0-\mu_0}{\sigma/\sqrt{n}}\right)=\alpha \\ &-z_{1-\alpha}=\frac{x_0-\mu_0}{\sigma/\sqrt{n}} \quad\Rightarrow\quad x_0=\mu_0-z_{1-\alpha}\cdot\sigma/\sqrt{n}.\end{aligned}$$

Best critical region:

$$C^\star=\left\{(x_1,\ldots,x_n):\overline{x}\le\mu_0-z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}\right\}\text{ in case of (a)}$$

$$C^\star=\left\{(x_1,\ldots,x_n):\overline{x}\ge\mu_0+z_{1-\alpha}\cdot\frac{\sigma}{\sqrt{n}}\right\}\text{ in case of (b).}$$

Generalised Likelihood Test

- This method can be used for composite hypotheses and multiple parameters

- Setting:
  - Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution
  - Denote $L(\widehat{\Omega}_0)$ the maximum of the likelihood function on the **restricted domain** $\Theta \in \Omega_0$, i.e.

$$L(\widehat{\Omega}_0) = \max_{\Theta \in \Omega_0} L(\Theta)$$

  - Similarly, let $L(\widehat{\Omega}) = \max_{\Theta \in \Omega} L(\Theta)$.
  - We define the likelihood ratio as

$$\Lambda = \frac{L(\widehat{\Omega}_0)}{L(\widehat{\Omega})} = \frac{\max_{\Theta \in \Omega_0} L(\Theta)}{\max_{\Theta \in \Omega} L(\Theta)}$$

  - Let k be a constant, the LRT of $H_0 : \Theta \in \Omega_0$ versus $H_1 : \Theta \in \Omega_1$ uses Λ as a test statistic with rejection region

$$\Lambda < k.$$

- The Test:
- Finding k:
  - Two situation can arise:
    1. We know the *exact* distribution of the test statistic Λ under H_0. In that case, we use this knowledge to find the precise k that makes the test of level α (as we did with the NPL).
    2. We do not know the distribution of Λ[1]. In that case, to find k we use an asymptotic result explained in the next slides.