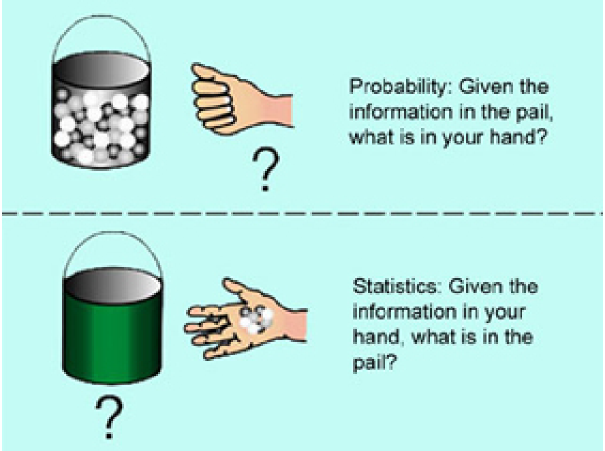
Lecture 5: Probability

Robert Garrard



Challange: Without using the phrases “chance” or “likelihood”, describe what a probability is.

Axioms of Probability

A **probability** is a function, $\mathbb{P} : 2^\Omega \to [0, 1]$ such that

1. $0 \leq \mathbb{P}(E) \leq 1$
2. $\mathbb{P}(\Omega) = 1$
3. For any countable sequence of *disjoint* events E_i,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$$

Probability *measures* the size of sets in a particular way.

NB: This is not entirely correct. The domain of the function is not the whole power set but a subset that is a "measurable space" and $\mathbb{P}$ must be a "measurable function", but this is beyond our scope.

Axioms of Probability

Exercise 5.1

1. Show that $\mathbb{P}(\emptyset) = 0$.
2. Show that $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.
3. Prove the inclusion-exclusion principle:
   $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
4. Two fair coins are flipped. The person flipping the coins reveals to you that one of the coins is a head. Find the probability that the other is tails.

Conditional Probability

Let A and B be events. The probability of A **conditional** on having observed B is

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

Two events A and B are said to be **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Exercise 5.2

Apply this formula to the previous coin-flipping example.

Law of Total Probability

Theorem (Law of Total Probability)

Consider an event A. Let $B_i, \quad i = 1, \ldots, n$ be a partition of the sample space. That is, $B_i \cap B_j = \emptyset \ \ \forall i, j$ and $\bigcup_{i=1}^{n} B_i = \Omega$.

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i)$$

It's often useful to substitute the conditional probability formula to make the theorem

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Law of Total Probability

Exercise 5.3

Consider two urns. The first contains two white and seven black balls, and the second contains five white and six black balls. We flip a fair coin and then draw a ball from the first urn or the second urn depending on whether the outcome was heads or tails.

What is the probability of drawing a white ball?

Bayes' Theorem

Bayes' theorem gives us a way to invert conditional probabilities.

Theorem

Bayes' Theorem Let A and B be events such that $\mathbb{P}(B) > 0$.

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

Bayes' theorem is often more easily applied by substituting the law of total probability into the denominator

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_i \mathbb{P}(B|A_i)\mathbb{P}(A_i)}$$

Bayes' Theorem

Exercise 5.4 A blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1 percent of the healthy persons tested. If 0.5 percent of the population actually has the disease, what is the probability a person has the disease given that his test result is positive?

Random Variables

It frequently occurs that in performing an experiment we are mainly interested in some functions of the outcome as opposed to the outcome itself.

These quantities of interest, or more formally, these real-valued functions defined on the sample space, are known as **random variables**.

Dice Rolling

Let $X : \Omega \to \mathbb{R}$, be the random variable defined by the sum of two fair dice.

$$
\begin{aligned}
&\mathbb{P}(X = 2) = \mathbb{P}\{(2, 2)\} = 1/36 \\
&\mathbb{P}(X = 3) = \mathbb{P}\{(1, 2), (2, 1)\} = 2/36 \\
&\mathbb{P}(X = 4) = \mathbb{P}\{(1, 3), (2, 2), (3, 1)\} = 3/36 \\
&\text{etc}
\end{aligned}
$$

Random Variables

The **cumulative distribution function** (cdf), $F(\cdot)$, of the random variable X is defined for any real number $b \in \mathbb{R}$, by

$$F(b) = \mathbb{P}(X \leq b)$$

Some properties of the cdf F are:

i) $F(b)$ is a nondecreasing function of b

ii) $\lim_{b \to \infty} F(b) = F(\infty) = 1$

iii) $\lim_{b \to -\infty} F(b) = F(-\infty) = 0$

All probability questions about X can be answered in terms of the cdf $F(\cdot)$.

Discrete Random Variables

When the image of a random variable, $X : \Omega \to \mathbb{R}$, is countable, we say that it is a **discrete** random variable.

For discrete rvs, we define the **probability mass function** (pmf) to be:

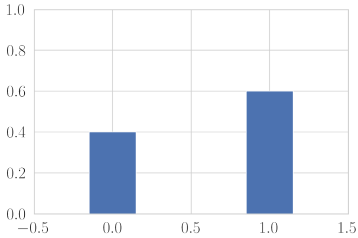
$$p_X(x_i) = \mathbb{P}(X = x_i)$$

This has the usual properties: $\forall x_i \; p_X(x_i) \geq 0$, and $\sum_i p_X(x_i) = 1$.

The pmf and cdf are related by:

$$p_X(x_i) = F(x_i) - F(x_{i-1}) \qquad F(x_i) = \sum_{j \leq i} p_X(x_j)$$

Common Discrete Random Variables



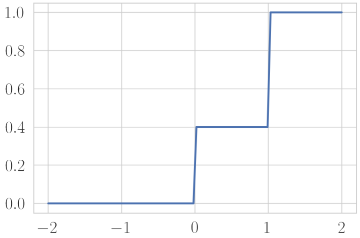


Figure: Bernoulli random variable

$$f(k;p) = \begin{cases} p & \text{if } k = 1 \\ 1-p & \text{if } k = 0 \end{cases} \qquad F(x;p) = \begin{cases} 0 & \text{if } x < 0 \\ 1-p & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

Common Discrete Random Variables

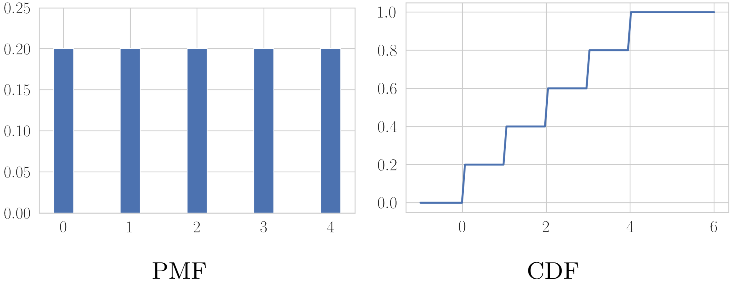
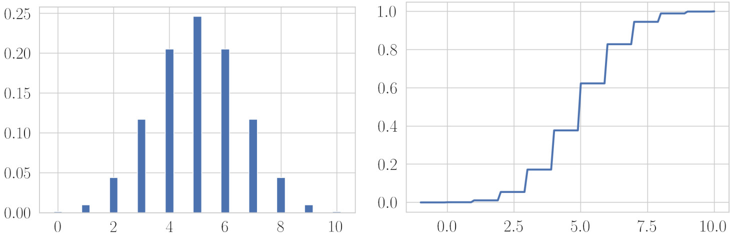


Figure: Discrete Uniform

$$f(k;a,b) = \begin{cases} \frac{1}{b-a+1} & \text{if } k \in \{a, a+1, \ldots, b\} \\ 0 & \text{Otherwise} \end{cases} \qquad F(x;a,b) = \frac{\lfloor k \rfloor - a + 1}{b-a+1}$$

Common Discrete Random Variables

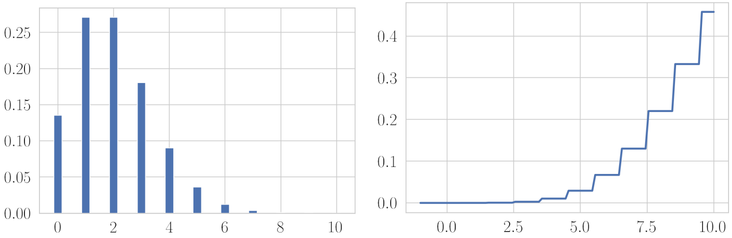


PMF CDF

Figure: Binomial

$$f(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Common Discrete Random Variables



PMF CDF

Figure: Poisson

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Common Discrete Random Variables

Exercise 5.5

Show that the poisson pmf is a valid pmf.

Continuous Random Variables

When the cdf of a random variable is continuous everywhere, we call it a **continuous** random variable.

If the cdf is **absolutely continuous** it admits a **probability density function** (pdf), $f(x)$, where:

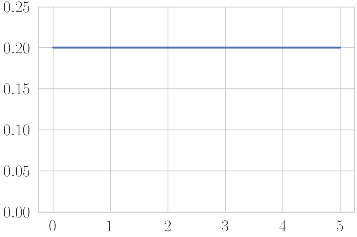
$$\forall x \ f(x) \geq 0 \qquad \int_{\infty}^{\infty} f(x) \mathrm{d}x = 1$$

$$F(b) = \int_{-\infty}^{b} f(x) \mathrm{d}x$$

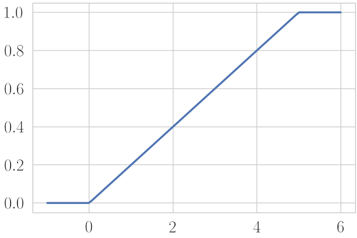
$$\mathbb{P}(a \leq X \leq b) = \int_{a}^{b} f(x) \mathrm{d}x$$

$$\frac{\mathrm{d}}{\mathrm{d}x} F(x) = f(x)$$

Common Continuous Random Variables



PDF

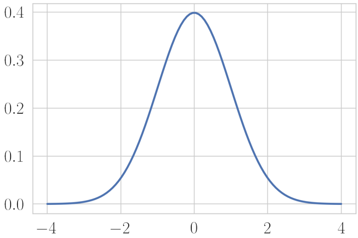


CDF

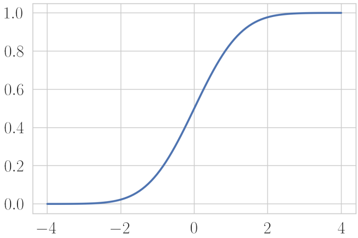
Figure: Discrete Uniform

$$f(x;a,b) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a,b) \\ 0 & \text{otherwise} \end{cases} \qquad F(x;a,b) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & x > b \end{cases}$$

Common Continuous Random Variables



PDF



CDF

Figure: Normal

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma^2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

Common Continuous Random Variables

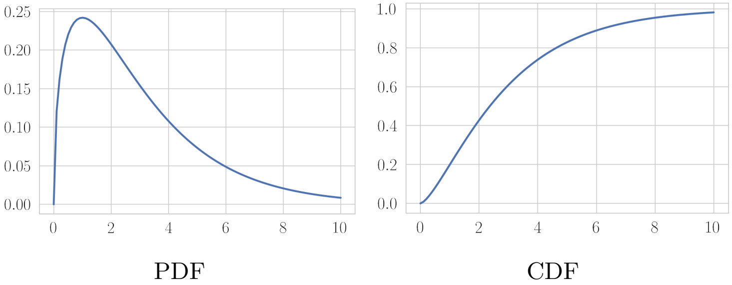


Figure: Chi-square

Let $Z_1, \ldots, Z_v$ be v independent standard normal distributions.

$$\sum_{i=1}^{v} Z_i^2 \sim \chi^2(v)$$

Common Continuous Random Variables

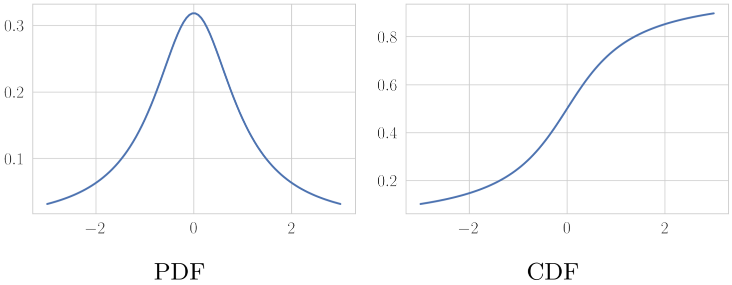


Figure: Student's t

Let $Z \sim N(0,1)$ and $V \sim \chi^2(v)$, Z and V independent.

$$\frac{Z}{\sqrt{V/v}} \sim t(v)$$

Common Continuous Random Variables

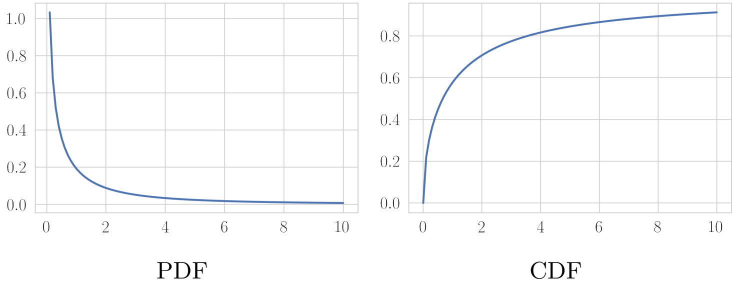


Figure: Fisher

Let $X \sim \chi^2(d_1)$ and $Y \sim \chi^2(d_2)$.

$$\frac{X/d_1}{Y/d_2} \sim F(d_1, d_2)$$

Expectation

Discrete Case:

If X is a discrete random variable having a probability mass function $\mathbb{P}(x)$, then the expected value of X is defined by

$$\mathbb{E}[X] = \sum_x x \cdot \mathbb{P}(x)$$

Continuous Case:

If X is a continuous random variable having a probability density function $f(x)$, then the expected value of X is defined by

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

For this to be well defined, we require that $\mathbb{E}[|X|] < \infty$.

Expectation

Theorem

(a) *If X is a discrete random variable with probability mass function $\mathbb{P}(x)$, then for any real-valued function g,*

$$\mathbb{E}[g(X)] = \sum_x g(x)\mathbb{P}(x)$$

(b) *If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g,*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$$

Expectation

Exercise 5.6

1. Show that if a and b are constants, $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.
2. I throw a dart uniformly at random at a dart board of unit radius. On average, what will be the distance my dart lands from the bullseye?

Economic Application: Decision Under Uncertainty

A simple **lottery** is a list $L = \{p_1, \dots, p_N\}$, with $p_n \geq 0$, $\sum_n p_n = 1$, where p_n is the interpreted as the probability of outcome n occurring.

We can form **compound lotteries**, $(L_1, \dots, L_k; \alpha_1, \dots, \alpha_k)$, that yields the simple lottery L_i with probability α_i.

Define a preference relation on the space of lotteries to be **continuous** if for any L, L', L'', the sets

$$\{\alpha \in [0,1] : \alpha L + (1-\alpha)L' \succsim L''\}$$

$$\{\alpha \in [0,1] : \alpha L + (1-\alpha)L' \precsim L''\}$$

are closed.

Economic Application: Decision Under Uncertainty

The preference relation satisfies the **independence axiom** if for all L, L', L''

$$L \succsim L' \iff \alpha L + (1-\alpha)L' \succsim \alpha L' + (1-\alpha)L''$$

Theorem

Expected Utility Theorem Suppose a rational preference relation, $\succeq$, on the set of lotteries satisfies the continuity and independence axioms. Then $\succeq$ can be represented in expected utility form

$$L \succeq L' \iff \mathbb{E}_L[U] \geq \mathbb{E}_{L'}[U]$$

for some utility function U.

Variance

The **variance** of a random variable X is defined as

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

often written more compactly as

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Exercise 5.7

1. Show that $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.

Moments

The k-th **central moment** is defined as

$$\mu_k = \mathbb{E}[(X - \mathbb{E}[X])^k]$$

The k-th **raw moment** is defined as

$$\mu'_k = \mathbb{E}[X^k]$$

If the nth moment exists, so does the $(n-1)$th.

Moment Generating Functions

If all of the moments exist, we can talk about the **moment generating function**.

$$M_X(t) = \mathbb{E}[e^{tX}]$$

The MGF has the following useful properties:

1. If random variables X and Y have the same MGF, they have the same distribution.
2. If X and Y are independent random variables, then their sum $X + Y$ has MGF: $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$.
3. $M_{cX}(t) = M_X(ct)$.
4. $\mu'_k = \frac{\mathrm{d}^n M_X}{\mathrm{d}t^n}\Big|_{t=0}$

Moment Generating Functions

Exercise 5.8

1. Show that the standard normal distribution has MGF: $M_Z(t) = e^{\frac{1}{2}t^2}$.
2. Let $X \sim N(\mu, \sigma^2)$. This has MGF $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$.

   Show that if $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent random variables, then if $Y = X_1 + X_2$ is the sum, $Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.
3. Show that if $X \sim N(\mu, \sigma^2)$, then $aX \sim N(a\mu, a^2\sigma^2)$.

Jointly Distributed Random Variables

Thus far we've concerned ourselves only with the probability distribution of a single random variable.

However, we are often interested in probability statements concerning two or more random variables.

To deal with such probabilities, we define, for any two random variables X and Y, the **joint cumulative distribution function** of X and Y by

$$F(a,b) = \mathbb{P}(X \leq a, Y \leq b) \qquad -\infty < a, b < \infty$$

Jointly Distributed Random Variables

The **marginal** distribution of X can be obtained from the joint distribution as follows

$$
\begin{aligned}
F_X(a) &= \mathbb{P}(X \leq a) \\
&= \mathbb{P}(X \leq a, Y \leq \infty) \\
&= F(a, \infty)
\end{aligned}
$$

Similarly for the CDF of Y.
As before, we'll talk about the **joint pmf** or **joint density**, $\mathbb{P}(x, y) = \mathbb{P}(X = x, Y = y)$. From this the marginal distributions are

$$
\mathbb{P}_X(x) = \sum_y \mathbb{P}(x, y) \qquad \mathbb{P}_X(x) = \int_{-\infty}^{\infty} \mathbb{P}(x, y)\mathrm{d}y
$$

Covariance

The **covariance** of any two random variables, X and Y, is

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

or sometimes more conveniently

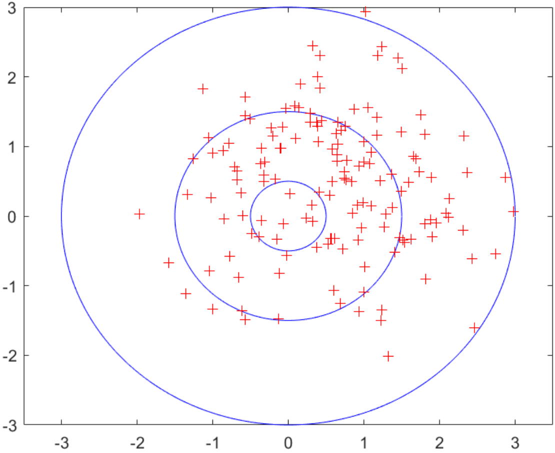
$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

We get **correlation** by scaling covariance by the appropriate variances.

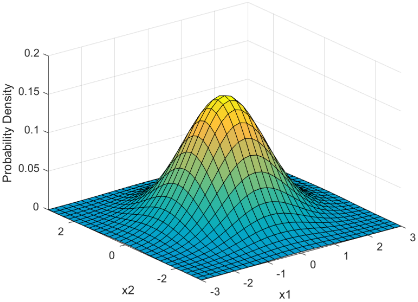
$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Exercise 5.9
Show that if X and Y are independent, then $\text{Cov}(X, Y) = 0$.

Random Vectors



Multivariate Normal



A random vector which is multivariate normally distributed with mean μ and covariance matrix $\mathbf{\Sigma}$, written $\mathbf{x} \sim N(\mu, \mathbf{\Sigma})$, has the following density function

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k |\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)\right)$$

Multivariate Normal

Proposition

Let $[X_1, \ldots, X_n]' \sim N(\mathbf{0}, \mathbf{\Sigma})$. If X_i and X_j are uncorrelated $i \neq j$, then X_i and X_j are independent.

Multivariate Normal

Proposition

Let $\mathbf{x} \sim N(\mu, \mathbf{\Sigma})$. Then for some non-zero matrix A (or vector) we have

$$\mathbf{Ax} \sim N(\mathbf{A}\mu, \mathbf{A\Sigma A}')$$

Exercise 5.10 Suppose $X \sim N(0,1)$ and $Y \sim N(0,1)$ such that $\mathrm{Cov}(X,Y) = 0.5$. What distribution does $X - Y$ have? Confirm this with a Monte Carlo experiment.

Distribution of Quadratic Forms

Proposition

Let $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$ be a $k \times 1$ MVN random vector.

$$\mathbf{x}'\mathbf{x} \sim \chi^2(k)$$

You can think about this as a quadratic form in the identity matrix: $\mathbf{x}'\mathbf{I}\mathbf{x}$.

Distribution of Quadratic Forms

Theorem

Let $\mathbf{x} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ *and let* M *be a symmetric idempotent matrix of rank* m. *Then*

$$\frac{\mathbf{x}' M \mathbf{x}}{\sigma^2} \sim \chi^2(m)$$

Exercise 5.11 Proof:

Inequalities

Theorem (Markov's Inequality)

If X is a random variable that takes only nonnegative values, then for any value $a > 0$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Proof:

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x f(x) \mathrm{d}x \\
&= \int_0^a x f(x) \mathrm{d}x + \int_a^\infty x f(x) \mathrm{d}x \\
&\geq \int_a^\infty x f(x) \mathrm{d}x \\
&\geq \int_a^\infty a f(x) \mathrm{d}x \\
&= a \mathbb{P}(X \geq a)
\end{aligned}$$

Inequalities

Theorem (Chebyshev's Inequality)

If X is a random variable with mean μ and variance σ^2, then for any $\varepsilon > 0$

$$\mathbb{P}(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Proof: Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov's inequality to obtain

$$\mathbb{P}\left((X - \mu)^2 \geq \varepsilon^2\right) \leq \frac{\mathbb{E}[(X - \mu)^2]}{\varepsilon^2}$$

But since $(X - \mu)^2 \geq \varepsilon^2$ iff $|X - \mu| > \varepsilon$,

$$\mathbb{P}(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Inequalities

Theorem (Jensen's Inequality)

Let X be a random variable and $f(\cdot)$ be a convex function. Then

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

Learning Outcomes

You should be able to:

- Calculate probabilities using the appropriate laws.
- Calculate expectations, variances, and covariances.
- Manipulate moment generating functions.
- Apply laws relating to random vectors.
- Apply inequalities where appropriate.