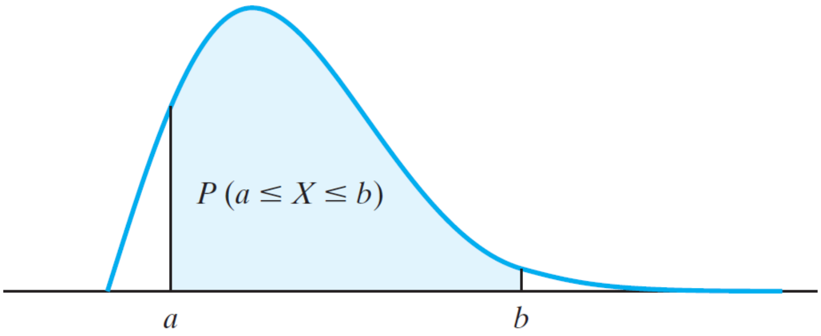
# CHAPTER 5: PROBABILITY DENSITIES

## 5.1 Continuous Random Variables

probability of random variable taking on a value in interval is given by:

$$P(a \leq X \leq b) = \sum_{i=1}^{m} f(x_i) \cdot \Delta x = \int_a^b f(x)dx$$




$$f(x) \geq 0 \; \forall x$$
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

**cumulative distribution function**

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

**mean of a probability density**

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

**kth moment about mean**

$$\mu_k = \int_{-\infty}^{\infty} (x-\mu)^k \cdot f(x)dx$$

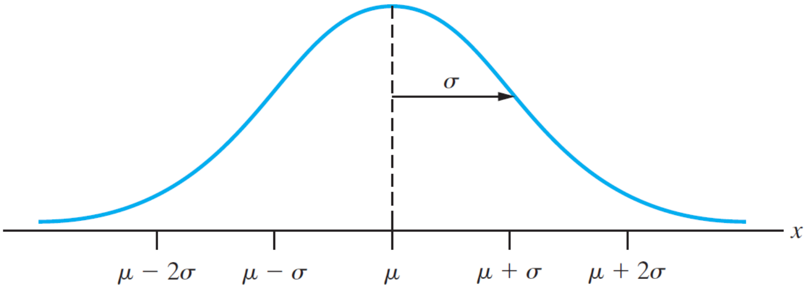
**variance of a probability density**

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 \\ &= E(X-\mu)^2 = E(X^2) - \mu^2 \end{aligned}$$

## 5.2 The Normal Distribution

normal probability density

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




**standard normal probabilities**

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt = P(Z \leq z)$$

**normal properties**

when X has normal distribution with mean $\mu$ and standard deviation $\sigma$

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

## 5.3 The Normal Approximation to the Binomial Distribution

**Normal approximation to binomial distribution**

If X is a random variable having the binomial distribution with parameters n & p, the limiting form of the distribution function of standardized random variable

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

as $n \to \infty$ is given by standard normal distribution

$$F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \qquad -\infty < z < \infty$$

rule: use only when $np$ and $n(1-p)$ are greater than 15

## 5.4 Other Probability Densities

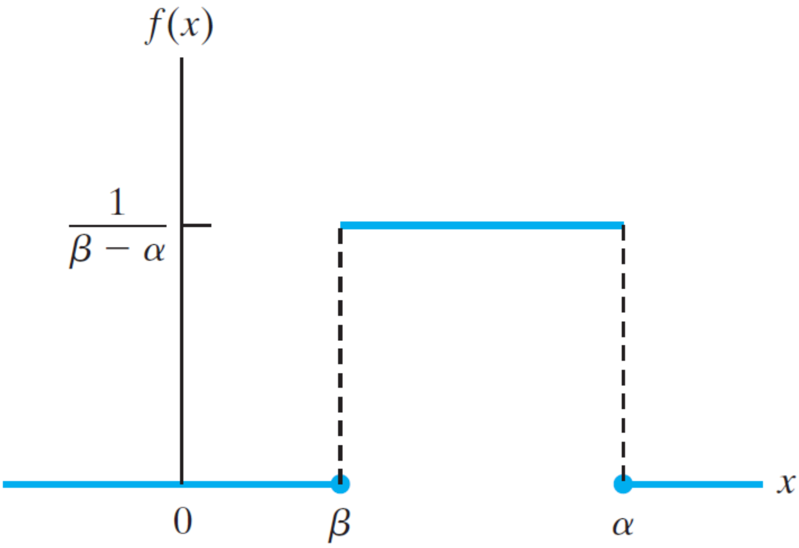
5 continuous distributions

- uniform distribution
- log-normal distribution
- gamma distribution
- beta distribution
- Weibull distribution

## 5.5 The Uniform Distribution

probability density function:

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{for } \alpha < x < \beta \\ 0 & \text{elsewhere} \end{cases}$$




mean of uniform distribution

$$\mu = \int_{\alpha}^{\beta} x \cdot \frac{1}{\beta - \alpha} dx = \frac{\alpha + \beta}{2}$$
$$\mu_2' = \int_{\alpha}^{\beta} x^2 \cdot \frac{1}{\beta - \alpha} dx = \frac{\alpha^2 + \alpha\beta + \beta^2}{3}$$

variance of uniform distribution

$$\sigma^2 = \mu_2' - \mu^2 = \frac{1}{12}(\beta - \alpha)^2$$

## 5.6 The Log-Normal Distribution

## 5.7 The Gamma Distribution

## 5.8 The Beta Distribution

## 5.9 The Weibull Distribution

## 5.10 Joint Distributions -- Discrete & Continuous

**Discrete Variables**

for 2 discrete variables $X_1, X_2$, and probability of values $x_1, x_2$, the probability of intersecting events & joint probability distribution:

$$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$$

**marginal probability distributions**

$$P(X_1 = x_1) = f_1(x_1) = \sum_{values\ x_2} f(x_1, x_2)$$

**conditional probability distribution**

$$f_1(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} \forall x_1 \text{ provided } f_2(x_2) \neq 0$$

if $f_1(x_1|x_2) = f_1(x_1) \forall x_1, x_2$, the conditional probability distribution is free of $x_2$, or equivalently if

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) \forall x_1, x_2$$

the 2 random variables are independent.

**Continuous Variables**

**joint probability density** $f(x_1, x_2, \ldots, x_k)$ of continuous random variables $X_1, X_2, \ldots X_k$ if probability that $a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \ldots, a_k \leq X_k \leq b_k$ is given by multiple integral

$$\int_{a_k}^{b_k} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2, \ldots, x_k) dx_1 dx_2 \ldots dx_k$$

**joint cumulative distribution function** F where kth random variable will take on value $\leq x_k$.

$f_i$: **marginal density** of ith random variable

$$f_i x_i = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_k) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_k$$

**Independent random variables**: k random variable $X_1, \ldots, X_k$ are independent iff

$$F(x_1, x_2, \ldots, x_k) = F_1(x_1) \cdot F_2(x_2) \ldots F_k(x_k)$$

for all values $x_1, x_2, \ldots, x_k$ of these random variables

**conditional probability density**

$$f_1(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} \text{ provided } f_2(x_2) \neq 0$$

**Properties of Expectation**

sum of the products of value x probability

**expected value of g(X)**:

- discrete: X has probability distribution $f(x)$
  $E[g(X)] = \sum_{x_i} g(x_i) f(x_i)$
- continuous: X has probability density function $f(x)$
  $E[g[X]] = \int_{-\infty}^{\infty} g(x) f(x) dx$

for given constants a and b

$$E(aX + b) = aE(X) + b$$
$$Var(aX + b) = a^2 Var(X)$$

**expected value of** $g(X_1, X_2, \ldots, X_k)$:

- discrete case
  $E[g(X_1, X_2, \ldots, X_k)] = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_k} g(x_1, x_2, \ldots, x_k) f(x_1, x_2, \ldots, x_k)$
- continuous case

$E[g(X_1, X_2, \ldots, X_k)]$
$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_k) f(x_1, x_2, \ldots, x_k) dx_1 dx_2 \ldots dx_k$

population **covariance**: measure $E[(X_1 - \mu_1)(X_2 - \mu_2)]$ of join variation

when $X_1$ and $X_2$ are independent, their covariance

$$E[(X_1 - \mu_1)(X_2 - \mu_2)] = 0$$

Further, the expectation of a linear combination of two independent random variables $Y = a_1 X_1 + a_2 X_2$ is

$$\begin{aligned}
\mu_Y = E(Y) &= E(a_1 X_1 + a_2 X_2) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a_1 x_1 + a_2 x_2) f_1(x) f_2(x_2)\, dx_1\, dx_2 \\
&= a_1 \int_{-\infty}^{\infty} x_1 f_1(x_1)\, dx_1 \int_{-\infty}^{\infty} f_2(x_2)\, dx_2 \\
&\quad + a_2 \int_{-\infty}^{\infty} f_1(x_1)\, dx_1 \int_{-\infty}^{\infty} x_2 f_2(x_2)\, dx_2 \\
&= a_1 E(X_1) + a_2 E(X_2)
\end{aligned}$$

This result holds even if the two random variables are not independent. Also,

$$\begin{aligned}
Var(Y) &= E(Y - \mu_Y)^2 = E[(a_1 X_1 + a_2 X_2 - a_1 \mu_1 - a_2 \mu_2)^2] \\
&= E[(a_1(X_1 - \mu_1) + a_2(X_2 - \mu_2))^2] \\
&= E[a_1^2 (X_1 - \mu_1)^2 + a_2^2 (X_2 - \mu_2)^2 + 2 a_1 a_2 (X_1 - \mu_1)(X_2 - \mu_2)] \\
&= a_1^2 E[(X_1 - \mu_1)^2] + a_2^2 E[(X_2 - \mu_2)^2] + 2 a_1 a_2 E[(X_1 - \mu_1)(X_2 - \mu_2)] \\
&= a_1^2\, Var(X_1) + a_2^2\, Var(X_2)
\end{aligned}$$

since the third term is zero because we assumed $X_1$ and $X_2$ are independent.

These properties hold for any number of random variables whether they are continuous or discrete.

mean & variance of linear combinations

Let $X_i$ have mean $\mu_i$ and variance $\sigma_i^2$ for $i = 1, 2, \ldots, k$. The linear combination $Y = a_1 X_1 + a_2 X_2 + \cdots + a_k X_k$ has

$$E(a_1 X_1 + a_2 X_2 + \cdots + a_k X_k) = a_1 E(X_1) + a_2 E(X_2) + \cdots + a_k E(X_k)$$

or

$$\mu_Y = \sum_{i=1}^{k} a_i \mu_i$$

When the random variables are independent,

$$\begin{aligned}
Var(a_1 X_1 + a_2 X_2 + \cdots + a_k X_k) = &\ a_1^2\, Var(X_1) \\
&+ a_2^2\, Var(X_2) + \cdots + a_k^2\, Var(X_k)
\end{aligned}$$

or

$$\sigma_Y^2 = \sum_{i=1}^{k} a_i^2 \sigma_i^2$$

## 5.11 Moment Generating Functions

## 5.12 Checking If the Data Are Normal

## 5.13 Transforming Observations to Near Normality