Mathematical Statistic Proof

Teng Man

October 31, 2022

Preliminaries

Math Stats (Mathematical Statistic) is hard because unlike previous math you learn are about applying, whereas the math stats is about writting and understanding proof. Yes, proof is hard becuase it is abstract and cumulative. In theoretical math, we starts from definitions and axioms, then deriving lemmas, theorems and corollaries.

Definition 0.1 (Expected Value)**.** Let X be a discrete random variable with the probability function $P(X = x)$. Then the *expected value* of X, $E(X)$, is defined to be

$$E(X) = \sum_{\text{all } x} xP(X = x)$$

1 Important Results

To understand the ideas behind the next theorem and its proof, we shall see a particular examples. A usual dice has faces of $1, 2, 3, 4, 5$ and 6. To describe the probability of dice-tossing, we may assume that the events of face facing upward are equally likely. Then, the distribution X can be represented by table below:

x	1	2	3	4	5	6
$P(X = x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

By definition 0.1, the expected value of this distribution is

$$\begin{aligned} E(X) &= \sum_{\text{all } x} xP(X = x) \\ &= 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) \\ &= \frac{7}{2} \end{aligned}$$

Suppose that a new dice having same faces as usual dice but with different number of dots on each face. The new faces are $0, 6, 0, 12, 0$ and 18. Indeed, these values are transformed by the function $y = f(x) = 3x$ if x is even else 0. Since having same number of faces, with the same assumption equally likely, the new distribution Y is similiar as before

y	0_1	6	0_3	12	0_5	18
$P(Y = y)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

y	0	6	12	18
$P(Y = y)$	$1/2$	$1/6$	$1/6$	$1/6$

It can be calculated that the new expected value is 6. Observe that the probability of 6 facing upward in new dice is same as 2 in usual dice. Also observe that the function f is not one-to-one. Nevertheless, the probabilities are corresponded between these two different distribution.

Theorem 1.1. Let X be a discrete random variable with the probability function $P(X = x)$ and $f(x)$ be a real-valued function of X. Then the *expected value* of $f(X)$, $E[f(X)]$, is given by

$$E[f(X)] = \sum_{\text{all } x} f(x)P(X = x)$$

Proof. Suppose that domain of X consists of countably infinite values, $x_1, x_2, \ldots$ The function f is arbitrary that we must consider the case of not one-to-one function. Suppose that $f(X) = Y$ (*i.e.* : image of f) consists of finite values, $y_1, \ldots, y_n$. Since the function f is arbitrary (*i.e.* : including the case of not one-to-one),

$$P(Y = y_i) = P(f(X) = y_i) = \sum_{\substack{\text{all } x \text{ s.t.} \\ f(x)=y}} P(X = x)$$

By definition 0.1, the expected value of $f(X)$ is given by

$$\begin{aligned}
E[f(X)] &= \sum_{\text{all } y} yP(Y = y) \\
&= \sum_{\text{all } y} y \sum_{\substack{\text{all } x \text{ s.t.} \\ f(x)=y}} P(X = x) \\
&= \sum_{\text{all y}} \sum_{\substack{\text{all } x \text{ s.t.} \\ f(x)=y}} yP(X = x) \\
&= \sum_{\text{all y}} \sum_{\substack{\text{all } x \text{ s.t.} \\ f(x)=y}} f(x)P(X = x) \\
&= \sum_{\text{all } x} f(x)P(X = x)
\end{aligned}$$

□

Usually, proofs start with suppositions (e.g. : any values, arbitrary values). Then, we can collect related definitions. For example, the previous proof mentions the definition of expected value. However, many texts omit the initial suppositions because these texts assume mature readers. Hence, you must include all necessary suppositions in writing rigorous proof.

Let X be any random variable (including discrete and continuous), c be arbitrary real number, and $f_1(x), \ldots, f_n(x)$ be any functions of X, then following are true

Theorem 1.2. $E(c) = c$

Theorem 1.3. $E[cf(x)] = cE[f(x)]$

Theorem 1.4. $E[f_1(x) + \ldots + f_n(x)] = E[f_1(x)] + \ldots + E[f_n(x)]$

Interested readers may consult the textbooks for the proofs of these theorems 1.2, 1.3 and 1.4.

Definition 1.1. Suppose X is any random variable, the variance of ta random variable X is defined as follow

$$V(X) = E\{[X - E(X)]^2\}$$

Theorem 1.5. If X is a random variable with probability function $P(X = x)$, then

$$V(X) = E\{[X - E(X)]^2\} = E(X^2) - [E(X)]^2$$

Proof. Let $E(X) = \mu$. Note that $E(X)$ is a constant value, hence

$$\begin{aligned}
V(X) &= E[(X-\mu)^2] && \text{(by definition of variance)} \\
&= E(X^2 - 2X\mu + \mu^2) \\
&= E(X^2) - 2E(X\mu) + E(\mu^2) && \text{(by theorem 1.4)} \\
&= E(X^2) - 2\mu E(X) + \mu^2 && \text{(by theorem 1.3)} \\
&= E(X^2) - 2\mu \cdot \mu + \mu^2 && \text{(by theorem 1.2)} \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - E(X)^2
\end{aligned}$$

□

Same as before, in writing proof, we can start with definitions after stating necessary suppositions. The theorems 1.2, 1.3 and 1.4 acts like lemmas (i.e.: helpers). They are used in proving the theorem 1.5. The proof demonstrates the usage of previous results. This shows the difficultly of math is due to its cumulative nature.

2 Binomial Distribution

Definition 2.1 (Binomial Coefficient)**.**

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

Theorem 2.1 (Binomial Theorem)**.** For any nonnegative integers n,

$$(x+y)^n = \sum_{k=0}^{n} x^k y^{n-k}$$

Definition 2.2 (Binomial Distribution)**.** A random variable X has a binomial distribution based on n trials with success probability p iff.

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ , } y = 0, 1, \ldots, n \text{ and } 0 \leq p \leq 1$$

By 2.1, it can be easily show that

Corollary 2.1.1. If X is a random binomial distribution and $q = 1 - p$, then

$$\sum_{i=0}^{n} P(X=x) = \sum_{i=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{i=0}^{n} \binom{n}{x} p^x q^{n-x} = 1$$

Proof. By Binomial Theorem,

$$\sum_{i=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1^n = 1$$

□

Theorem 2.2 (mean and variance of binomial distribution)**.** If X is a random binomial distribution based on n trials with success probability p, then the expected value and variance of X are

$$E(X) = np \text{ and } V(X) = np(1-p)$$

Proof. Let $q = 1-p$. By definition of Expected Value and Binomial Distribution,

$$E(X) = \sum_{\text{all } x} xP(X=x) = \sum_{i=0}^{n} x \binom{n}{x} p^x q^{n-x}$$

Note that the first term of x is 0, hence

$$\begin{aligned}
\sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} &= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{x(x-1)(n-x)!} p^x q^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x}
\end{aligned}$$

It is possibly easier if we can use the axiom stating that $\sum_{\text{all } x} P(X = x) = 1$. Refering the corollary 2.1.1, we may reorganize the summands,

$$\begin{aligned}
\sum_{x=0}^{n} x\binom{n}{x}p^x q^{n-x} &= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!}p^x q^{n-x} \\
&= \sum_{x=1}^{n} \frac{n(n-1)!}{(x-1)!(n-x)!}p \cdot p^{x-1} q^{n-x} \\
&= np\sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!}p^{x-1} q^{n-x}
\end{aligned}$$

Let $z = x - 1$ and $m = n - 1$, then $x = z + 1$, $x = 1 \implies z = 0$, $x = n \implies z = n - 1$ and

$$\begin{aligned}
\sum_{x=0}^{n} x\binom{n}{x}p^x q^{n-x} &= np\sum_{z=0}^{n-1} \frac{(n-1)!}{z!(n-1-z)!}p^z q^{n-1-z} \\
&= np\sum_{z=0}^{m} \frac{m!}{z!(m-z)!}p^z q^{m-z} \\
&= np\sum_{z=0}^{m} \binom{m}{z}p^z q^{m-z} \\
E(X) &= np
\end{aligned}$$

where $\sum_{z=0}^{m}\binom{m}{z}p^z q^{m-z}$ is resemblance to new binomial distribution, hence it follows from the corollary 2.1.1 that it is 1.

From the theorem 1.5, we know that $V(X) = E(X^2) - [E(X)]^2$. Thus, we can obtain the $V(X)$ if we obtain $E(X^2)$. Using previous technique doesn't work for finding $E(X^2)$,

$$\begin{aligned}
E(X^2) &= \sum_{x=0}^{n} x^2 P(X = x) = \sum_{x=0}^{n} x^2\binom{n}{x}p^x q^{n-x} \\
&= \sum_{x=0}^{n} x^2\frac{n!}{x!(n-1)!}p^z q^{m-z} \\
&= \sum_{x=0}^{n} x\frac{n!}{(x-1)!(n-1)!}p^z q^{m-z}
\end{aligned}$$

Instead, we include $x - 1$ so that it can be factored out. In fact,

$$\begin{aligned}
E[X(X-1)] &= E(X^2 - X) = E(X^2) - E(X) \\
\therefore E(X^2) &= E[X(X-1)] + E(X)
\end{aligned}$$

Note that the first 2 terms of x are 0 and 1, then their summands are zero. Thus

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{n} x(x-1)P(X=x) \\
&= \sum_{x=2}^{n} x(x-1)\binom{n}{x}p^x q^{n-x} \\
&= \sum_{x=2}^{n} x(x-1)\frac{n!}{x(x-1)(x-2)!n!}p^n q^{n-x} \\
&= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-1)!}p^n q^{n-x} \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!n!}p^{n-2} q^{n-x}
\end{aligned}$$

Using similar techinque in derivation of mean. Let $z = x - 2$, then

$$\begin{aligned}
E[X(X-1)] &= n(n-1)p^2 \sum_{z=0}^{n-2} \frac{(n-2)!}{z!n!}p^z q^{n-2-z} \\
&= n(n-1)p^2 \sum_{z=0}^{n-2} \binom{n-2}{z}p^z q^{n-2-z} \\
&= n(n-1)p^2 = n^2p^2 - np^2 \\
E[X(X-1)] &= (np)^2 - np^2
\end{aligned}$$

Therefore,

$$\begin{aligned}
V(X) &= E(X^2) - [E(X)]^2 \\
&= E[X(X-1)] + E(X) - [E(X)]^2 \\
&= (np)^2 - np^2 + np - (np)^2 \\
&= np - np^2 = np(1-p)
\end{aligned}$$

□

3 Geometric Distribution

Definition 3.1 (Geometric Distribution)**.** A random variable X has a geometric probability distribution iff.

$$P(X = x) = (1-p)^{x-1}p, x = 1, 2, 3, \ldots, 0 \leq p \leq 1$$

Theorem 3.1 (mean and variance of geometric distribution)**.** If X is a random geometric distribution with success probability p, then

$$E(X) = \frac{1}{p} \text{ and } V(X) = \frac{1-p}{p^2}$$

Proof. By definition of Expected Value, let $q = 1-p$ then $p = 1-q$ and

$$E(X) = \sum_{x=1}^{\infty} xq^{x-1}p = p\sum_{x=1}^{\infty} xq^{x-1}$$

Observe that, for $x \geq 1$

$$\int q^{x-1}dq = \frac{q^x}{x} \iff \frac{d}{dq}q^x = xq^{x-1}$$

Then,

$$E(X) = p\sum_{x=1}^{\infty} xq^{x-1} = p\frac{d}{dq}\left(\sum_{x=1}^{\infty} q^x\right)$$

The latter sum is the geometric series, $q + q^2 + q^3 + \ldots$, and if $|q| < 1$, then the sum is equal to $\frac{q}{1-q}$. Therefore,

$$\begin{aligned}
E(X) &= p\frac{d}{dq}\left(\frac{q}{1-q}\right) = p\frac{d}{dq}\left[q(1-q)^{-1}\right] \\
&= p\left[1 \cdot (1-q)^{-1} + q \cdot (-1 \cdot -1)(1-q)^{-2}\right] \\
&= p\left[\frac{1}{1-q} + \frac{q}{(1-q)^2}\right] \\
&= p\left[\frac{(1-q)+q}{(1-q)^2}\right] = p\left(\frac{1}{p^2}\right) \\
E(X) &= \frac{1}{p}
\end{aligned}$$

In deriving variance, we can use the same technique in deriving mean. Observe that,

$$\iint q^{x-1}dq = \frac{q^{x+1}}{x(x+1)} \iff \frac{d^2}{dq^2}q^{x+1} = x(x+1)q^{x-1}$$

Hence, the tedious second derivative calculation is omitted here and left as exercise for reader

$$
\begin{aligned}
E[X(X+1)] &= \sum_{x=1}^{\infty} x(x+1)q^{x-1}p = p\sum_{x=1}^{\infty} x(x+1)q^{x-1} \\
E(X^2+X) &= p\frac{d^2}{dq^2}\left(\sum_{x=1}^{\infty} q^{x+1}\right) \\
E(X^2)+E(X) &= p\frac{2}{(1+q)^3} \\
E(X^2)+\frac{1}{p} &= \frac{2p}{p^3} \\
E(X^2) &= \frac{2}{p^2}-\frac{1}{p}
\end{aligned}
$$

Then,

$$
\begin{aligned}
V(X) &= E(X^2)-[E(X)]^2 \\
&= \frac{2}{p^2}-\frac{1}{p}-\frac{1}{p^2} \\
&= \frac{1}{p^2}-\frac{1}{p} \\
V(X) &= \frac{1-p}{p^2}
\end{aligned}
$$

□

4 Poisson Distribution

Theorem 4.1. From calculus,

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Definition 4.1. A random variable X has a *Poisson probability distribution* iff.

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

Corollary 4.1.1. For poisson random variable X, the following is true

$$\sum_{\text{all } x} P(X = x) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} = 1$$

Proof.

$$\begin{aligned}
\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\
&= e^{-\lambda} \cdot e^{\lambda} \\
&= 1
\end{aligned}$$

□

Theorem 4.2 (mean and variance of poisson distribution)**.** If X is a random poisson distribution with parameter λ, then

$$E(X) = \lambda \text{ and } V(X) = \lambda$$

Proof. Note that the first term of the summands is 0,

$$\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} = \sum_{x=1}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} \\
&= \sum_{x=1}^{\infty} x \frac{\lambda \lambda^{x-1} e^{-\lambda}}{x(x-1)!} \\
&= \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!}
\end{aligned}$$

Let $z = x - 1$, then

$$E(X) = \lambda \sum_{z=0}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!} = \lambda$$

The latter summation is geometric distribution sum as in , then it is 1.

Note that the first 2 terms of the summands are 0,

$$
\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1)\frac{\lambda^x e^{-\lambda}}{x!} \\
E(X^2 - X) &= \sum_{x=2}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-2)!} \\
E(X^2) - E(X) &= \lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2} e^{-\lambda}}{(x-2)!} \\
\text{Let z = x - 2} \\
E(X^2) - \lambda &= \lambda^2 \sum_{z=0}^{\infty} \frac{\lambda^z e^{-\lambda}}{z!} = \lambda^2(1) \\
E(X^2) &= \lambda^2 + \lambda \\
\therefore V(X) &= E(X^2) - [E(X)^2] \\
&= \lambda^2 + \lambda - (\lambda)^2 \\
V(X) &= \lambda
\end{aligned}
$$

□

5 Continuous Random Variable

Definition 5.1. Let X denote any random variable. The *distribution function* of X, denoted by $F(x)$ is such that $F(x) = P(X \leq x)$ for $-\infty < x < \infty$. And X is continuous random variable iff. $F(X)$ is continuous for $-\infty < x < \infty$

Definition 5.2. Let X denote any continuous random variable with distribution function $F(x)$. If this derivative exists

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

, then $f(x)$ is called the *probability density function* for the continuous random variable X

Corollary. It follows from the definitions that

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

Theorem 5.1. If the random variable X with the density function $f(x)$ and $a < b$, then

$$P(a \leq x \leq b) = \int_{a}^{b} f(x)dx$$

Definition 5.3 (Expected Value)**.** Let X be a continuous random variable with the density function $f(x)$. Then the *expected value* of X, $E(X)$, is defined to be

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

6 Uniform Probability distribution

Definition 6.1. For $\theta_1 < \theta_2$, a random vairable X has a continuous *uniform distribution* on the interval (θ_1, θ_2) iff. the density function of X is

$$f(x) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & , \theta_1 < x < \theta_2, \\ 0 & , \text{elsewhere.} \end{cases}$$

Theorem 6.1 (mean and variance of uniform distribution)**.** For $\theta_1 < \theta_2$ and X is a continuous random variable on the interval (θ_1, θ_2), then

$$E(X) = \frac{\theta_1 + \theta_2}{2} \text{ and } V(X) = \frac{(\theta_2 - \theta_1)^2}{12}$$

Proof. The proof is left as exercises for readers. (common phrases in math textbooks) □

7 Gamma Probability Distribution

Definition 7.1 (Gamma Function)**.**

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

Theorem 7.1 (integration by part)**.**

$$\int udv = uv - \int vdu$$

Theorem 7.2. For n is any integer larger than 0

$$\Gamma(1) = 1$$

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha)$$

$$\Gamma(n) = (n-1)!$$

Proof.

$$\begin{aligned}
\Gamma(1) &= \int_0^\infty x^{1-1} e^{-x} dx = \int_0^\infty e^{-x} dx \\
&= \left[-e^{-x}\right]_0^\infty \\
&= -\lim_{x\to\infty} \frac{1}{e^x} - (-e^{-0}) \\
&= -0 - (-1) = 1
\end{aligned}$$

Integration by part,

$$\begin{aligned}
\Gamma(\alpha) &= \int_0^\infty x^{\alpha-1} e^{-x} dx \\
u &= x^{\alpha-1} & v &= e^{-x} \\
du &= (\alpha-1)x^{\alpha-2} dx & dv &= -e^{-x} dx \\
\Gamma(\alpha) &= \left[x^{\alpha-1}e^{-x}\right]_0^\infty - \int_0^\infty -e^{-x}(\alpha-1)x^{\alpha-2} dx \\
&= 0 + (\alpha-1)\int_0^\infty x^{\alpha-2} e^{-x} dx \\
\Gamma(\alpha) &= (\alpha-1)\Gamma(\alpha-2)
\end{aligned}$$

The third statement follows from the mathematical induction and $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha)$ □

Definition 7.2. A random variable X has a *gamma distribution with parameters* $\alpha > 0$ and $\beta > 0$ iff. the density function of X is

$$f(x) = \begin{cases} \frac{x^{\alpha-1} e^{\frac{-x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} & , 0 \le x < \infty, \\ 0 & , \text{elsewhere} \end{cases}$$

Theorem 7.3 (mean and variance of gamma distribution)**.** If a random variable X has a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$, then

$$E(X) = \alpha\beta \text{ and } V(X) = \alpha\beta^2$$

Proof. Since it is a probability distribution, by definition(i.e.: suppose for magic), the gamma density function is

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^{\infty} \frac{x^{\alpha-1}e^{\frac{-x}{\beta}}}{\beta^\alpha\Gamma(\alpha)}dx = 1$$

Hence,

$$\beta^\alpha\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{\frac{-x}{\beta}}\,dx$$

and

$$\begin{aligned}
E(X) &= \int_0^{\infty} x\frac{x^{\alpha-1}e^{\frac{-x}{\beta}}}{\beta^\alpha\Gamma(\alpha)}dx = \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^{\infty} x^\alpha e^{\frac{-x}{\beta}}\,dx \\
&= \frac{1}{\beta^\alpha\Gamma(\alpha)}\left[\beta^{\alpha+1}\cdot\Gamma(\alpha+1)\right] \\
&= \frac{1}{\Gamma(\alpha)}\left[\beta\cdot\alpha\Gamma(\alpha)\right] \\
E(X) &= \alpha\beta
\end{aligned}$$

Since $V(X) = E(X^2) - [E(X)]^2$,

$$\begin{aligned}
E(X^2) &= \int_0^{\infty} x^2\frac{x^{\alpha-1}e^{\frac{-x}{\beta}}}{\beta^\alpha\Gamma(\alpha)}dx = \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^{\infty} x^{\alpha+1}e^{\frac{-x}{\beta}}\,dx \\
&= \frac{1}{\beta^\alpha\Gamma(\alpha)}\left[\beta^{\alpha+2}\cdot\Gamma(\alpha+2)\right] \\
&= \frac{1}{\Gamma(\alpha)}\left[\beta^2\cdot(\alpha+1)\cdot\alpha\Gamma(\alpha)\right] \\
E(X^2) &= \alpha(\alpha+1)\beta^2
\end{aligned}$$

Then,

$$\begin{aligned}
V(X) &= E(X^2) - [E(X)]^2 \\
&= \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 \\
&= (\alpha^2+\alpha)\beta^2 - \alpha^2\beta^2 \\
&= (\alpha^2+\alpha-\alpha^2)\beta^2 \\
V(X) &= \alpha\beta^2
\end{aligned}$$

□

Definition 7.3 (Beta Function)**.**

$$B(\alpha, \beta) = \int^1 0x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Definition 7.4. A random variable X has a *beta distribution with parameters* $\alpha > 0$ and $\beta > 0$ iff. the density function of X is

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

where

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Theorem 7.4 (mean and variance of beta distribution)**.** If a random variable X has a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$, then

$$E(X) = \frac{\alpha}{\alpha+\beta} \text{ and } V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Theorem 7.5. By definition Expected Value,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_0^1 x \cdot \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx \\ &= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1}dx \\ &= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} \\ &= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \times \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\ &= \frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} \times \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\ E(X) &= \frac{\alpha}{\alpha+\beta} \end{aligned}$$

Since variance $V(X) = E(X^2) - [E(X)]^2$,

$$
\begin{aligned}
E(X^2) &= \int_0^1 x^2 \cdot \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} dx \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha+1}(1-x)^{\beta-1} dx \\
&= \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)} \\
&= \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \times \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\
&= \frac{\alpha(\alpha+1)\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)} \times \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\
E(X^2) &= \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
V(X) &= E(X^2) - [E(X)]^2 \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
&= \frac{(\alpha^2+\alpha)(\alpha+\beta) - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta+1)(\alpha+\beta)^2} \\
&= \frac{\alpha^2(\alpha+\beta) + \alpha(\alpha+\beta) - \alpha^2(\alpha+\beta) - \alpha^2}{(\alpha+\beta+1)(\alpha+\beta)^2} \\
&= \frac{\alpha^2 + \alpha\beta - \alpha^2}{(\alpha+\beta+1)(\alpha+\beta)^2} \\
&= \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}
\end{aligned}
$$