and in particular,

$$\mathbb{E}X = \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha}{\alpha+\beta}$$

and

$$\mathbb{E}X^2 = \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}.$$

Then

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$$

We now study important inequalities in connection with moments of RVs and probabilities of events involving the RVs. Given any RV X, we shall always assume that it is either discrete with p.m.f. f_X or continuous with p.d.f. f_X, if not stated otherwise.

Note 1.252. At times, it is possible to compute the moments of an RV, but the computation of probability of certain events involving the RV may be difficult. The inequalities, that we are going to study, give us estimates of the probabilities in question.

Theorem 1.253. *Let X be an RV such that X is non-negative (i.e. $\mathbb{P}(X \geq 0) = 1$). Suppose that $\mathbb{E}X$ exists. Then for any $c > 0$, we have*

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}X}{c}.$$

Proof. We discuss the proof when X is a continuous RV with p.d.f. f_X. The case when X is discrete can be proved using similar arguments.
For $x < 0$, we have $F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X < 0) = 1 - \mathbb{P}(X \geq 0) = 0$ and hence $f_X(x) = 0, \forall x < 0$. Then,

$$\mathbb{E}X = \int_0^\infty x f_X(x)\,dx \geq \int_c^\infty x f_X(x)\,dx \geq c\int_c^\infty f_X(x)\,dx = c\,\mathbb{P}(X \geq c).$$

This completes the proof. $\square$

Note 1.254. Under the assumptions of Theorem 1.253, we have $\mathbb{E}X \geq 0$.

The following special cases of Theorem 1.253 are quite useful in practice.

Corollary 1.255. *(a) Let X be an RV and let $h : \mathbb{R} \to [0, \infty)$ be a function such that $\mathbb{E}h(X)$ exists. Then for any $c > 0$, we have*

$$\mathbb{P}(h(X) \geq c) \leq \frac{\mathbb{E}h(X)}{c}.$$

(b) Let X be an RV and let $h : \mathbb{R} \to [0, \infty)$ be a strictly increasing function such that $\mathbb{E}h(X)$ exists. Then for any $c > 0$, we have

$$\mathbb{P}(X \geq c) = \mathbb{P}(h(X) \geq h(c)) \leq \frac{\mathbb{E}h(X)}{h(c)}.$$

(c) Let X be an RV such that $\mathbb{E}X$ exists, i.e. $\mathbb{E}|X| < \infty$. Considering the RV $|X|$, for any $c > 0$ we have

$$\mathbb{P}(|X| \geq c) \leq \frac{\mathbb{E}|X|}{c}.$$

(d) (Markov's inequality) Let $r > 0$ and let X be an RV such that $\mathbb{E}|X|^r < \infty$. Then for any $c > 0$, we have

$$\mathbb{P}(|X| \geq c) = \mathbb{P}(|X|^r \geq c^r) \leq c^{-r}\, \mathbb{E}|X|^r.$$

(e) (Chernoff's inequality) Let X be an RV with $\mathbb{E}e^{\lambda X} < \infty$ for some $\lambda > 0$. Then for any $c > 0$, we have

$$\mathbb{P}\{X \geq c\} = \mathbb{P}\{e^{\lambda X} \geq e^{\lambda c}\} \leq e^{-\lambda c}\, \mathbb{E}e^{\lambda X}.$$

Note 1.256. Let X be an RV with finite second moment, i.e. $\mu_2' = \mathbb{E}X^2 < \infty$. By Remark 1.201, the first moment $\mu_1' = \mathbb{E}X$ exists. Hence

$$\mathbb{E}(X - c)^2 = \mathbb{E}[X^2 + c^2 - 2cX] = \mathbb{E}X^2 + c^2 - 2c\,\mathbb{E}X = \mu_2' + c^2 - 2c\,\mu_1' < \infty$$

Therefore, all second moments of X about any point $c \in \mathbb{R}$ exists. In particular, $Var(X) = \mathbb{E}(X - \mu_1')^2 < \infty$. By a similar argument, for any RV X with finite variance, we have $\mathbb{E}X^2 < \infty$.

The next result is a special case of Markov's inequality.

Corollary 1.257 (Chebyshev's inequality)**.** *Let X be an RV with finite second moment (equivalently, finite variance). Then*

$$\mathbb{P}[|X-\mu_1'| \geq c] \leq \frac{1}{c^2}\mathbb{E}(X-\mu_1')^2 = \frac{1}{c^2}Var(X).$$

Remark 1.258. Another form of the above result is also useful. Under the same assumptions, for any $\epsilon > 0$ we have

$$\mathbb{P}[|X-\mu_1'| \geq \epsilon\,\sigma(X)] \leq \frac{1}{\epsilon^2},$$

where $\sigma(X)$ is the standard deviation of X. This measures the spread/deviation of the distribution (of X) about the mean in multiples of the standard deviation. The smaller the variance, lesser the spread.

Remark 1.259. In general, bounds in Theorem 1.253 or in Markov/Chebyshev's inequalities are very conservative. However, they can not be improved further. To see this, consider a discrete RV X with p.m.f. given by

$$f_X(x) := \begin{cases} \frac{3}{4}, & \text{if } x = 0, \\ \frac{1}{4}, & \text{if } x = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then $\mathbb{P}(X \geq 1) = \frac{1}{4} = \mathbb{E}X$, which is sharp. If we consider

$$f_X(x) := \begin{cases} \frac{3}{4}, & \text{if } x = 0, \\ \frac{1}{4}, & \text{if } x = 2, \\ 0, & \text{otherwise,} \end{cases}$$

then, $\mathbb{P}(X \geq 1) = \frac{1}{4} < \frac{1}{2} = \mathbb{E}X$.

Definition 1.260 (Convex functions)**.** Let I be an open interval in $\mathbb{R}$. We say that a function $h : I \to \mathbb{R}$ is convex on I if

$$h(\alpha x + (1-\alpha)y) \leq \alpha h(x) + (1-\alpha)h(y), \forall \alpha \in (0,1), \forall x, y \in I.$$

We say that h is strictly convex on I if the above inequality is strict for all x, y and α.

We state the following result from Real Analysis without proof.

Theorem 1.261. *Let I be an open interval in $\mathbb{R}$ and let $h : I \to \mathbb{R}$ be a function.*

(a) If h is convex on $\mathbb{R}$, then h is continuous on $\mathbb{R}$.

(b) Let h be twice differentiable on I. Then,

  (i) h is convex if and only if $h''(x) \geq 0, \forall x \in I$.

  (ii) h is strictly convex if and only if $h''(x) > 0, \forall x \in I$.

The following result is stated without proof.

Theorem 1.262 (Jensen's Inequality)**.** *Let I be an interval in $\mathbb{R}$ and let $h : I \to \mathbb{R}$ be a convex function. Let X be an RV with support $S_X \subseteq I$. Then,*

$$h(\mathbb{E}X) \leq \mathbb{E}h(X),$$

provided the expectations exist. If h is strictly convex, then the inequality above is strict unless X is a degenerate RV.

Remark 1.263. Some special cases of Jensen's inequality are of interest.

(a) Consider $h(x) = x^2, \forall x \in \mathbb{R}$. Here, $h''(x) = 2 > 0, \forall x$ and hence h is convex on $\mathbb{R}$. Then $(\mathbb{E}X)^2 \leq \mathbb{E}X^2$, provided the expectations exist. We had seen this inequality earlier in Remark 1.204.

(b) For any integer $n \geq 2$, consider the function $h(x) = x^n$ on $[0, \infty)$. Here, $h''(x) = n(n-1)x^{n-2} \geq 0, \forall x \in (0, \infty)$ and hence h is convex. Then $(\mathbb{E}|X|)^n \leq \mathbb{E}|X|^n$, provided the expectations exist.

(c) Consider $h(x) = e^x, \forall x \in \mathbb{R}$. Here, $h''(x) = e^x > 0, \forall x$ and hence h is convex on $\mathbb{R}$. Then $e^{\mathbb{E}X} \leq \mathbb{E}e^X$, provided the expectations exist.

(d) Consider any RV X with $\mathbb{P}(X > 0) = 1$ and look at $h(x) := -\ln x, \forall x \in (0, \infty)$. Then $h''(x) = \frac{1}{x^2} > 0, \forall x \in (0, \infty)$ and hence h is convex. Then $-\ln(\mathbb{E}X) \leq \mathbb{E}(-\ln X)$, i.e. $\ln(\mathbb{E}X) \geq \mathbb{E}(\ln X)$, provided the expectations exist.

(e) Consider any RV X with $\mathbb{P}(X > 0) = 1$. Then $\mathbb{P}(\frac{1}{X} > 0) = 1$ and hence by (d), $-\ln(\mathbb{E}\frac{1}{X}) \leq \mathbb{E}(-\ln \frac{1}{X}) = \mathbb{E}(\ln X)$. Then $(\mathbb{E}\frac{1}{X})^{-1} = e^{-\ln(\mathbb{E}\frac{1}{X})} \leq e^{\mathbb{E}(\ln X)} \leq \mathbb{E}X$, by (c). This inequality holds, provided all the expectations exist. We may think of $\mathbb{E}X$ as the arithmetic mean (A.M.) of X, $e^{\mathbb{E}(\ln X)}$ as the geometric mean (G.M.) of X, and $\frac{1}{\mathbb{E}[\frac{1}{X}]}$ as the harmonic mean (H.M.) of X. The inequality obtained here is related to the classical A.M.-G.M.-H.M. inequality (see problem set 6).

Note 1.264 (Why should we look at multiple RVs together?)**.** Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ associated with a random experiment $\mathcal{E}$. As motivated earlier, an RV associates some numerical quantity to each of the outcomes of the experiment. Such numerical quantities help us in the understanding of characteristics of the outcomes. However, it is important to note that, in practice, we may be interested in looking at these characteristics of the outcomes at the same time. This also allows us to see if the characteristics in question may be related. If we perform the random experiment separately for each of these characteristics, then there is also the issue of cost and time associated with the repeated performance of the experiment. Keeping this in mind, we now choose to consider multiple characteristics of the outcomes at the same time. This leads us to the concept of Random Vectors, which allows us to look at multiple RVs at the same time.

Example 1.265. Consider the random experiment of rolling a standard six-sided die three times. Here, the sample space is

$$\Omega = \{(i, j, k) : i, j, k \in \{1, 2, 3, 4, 5, 6\}\}.$$

Suppose we are interested in the sum of the first two rolls and the sum of all rolls. These characteristics of the outcomes can be captured by the RVs $X, Y : \Omega \to \mathbb{R}$ defined by $X((i, j, k)) := i + j$ and $Y((i, j, k)) := i + j + k$ for all $(i, j, k) \in \Omega$. If we look at X and Y simultaneously, we may comment on whether a 'large' value for X implies a 'large' Y and vice versa.

Definition 1.266 (Random Vector)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X = (X_1, X_2, \cdots, X_p) : \Omega \to \mathbb{R}^p$ is called a p-dimensional random vector (or simply, a random vector, if the dimension p is clear from the context). Here, the component functions are denoted by $X_1, X_2, \cdots, X_p$ and each of these are real valued functions defined on the sample space Ω and hence are RVs.