Chapter Seven
Properties of Expectation

**Proposition 2.1**
If $X$ and $Y$ have a joint probability mass function $p(x,y)$, then

$$E[g(X,Y)] = \sum_y \sum_x g(x,y) p(x,y),$$

provided that the sum is absolutely convergent.

If $X$ and $Y$ have a joint probability density function $f(x,y)$, then

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dx dy,$$

provided that the integral is absolutely convergent.

**Example 2a.** An accident occurs at a point $X$ that is uniformly distributed on a road of length $L$. At the time of the accident an ambulance is at a location $Y$ that is also uniformly distributed on the road. Assuming that $X$ and $Y$ are independent, find the expected distance between the ambulance and the point of the accident. Answer is $L/3$.

A corollary of this theorem is that $E[X + Y] = E[X] + E[Y]$ provided that the expectations are finite. By induction, $E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$, provided each expectation is finite.

**Example 2c.** Let $X_1, X_2, \ldots, X_n$ be jointly distributed random variables, each with expected value $\mu$. Prove that $E[\overline{X_n}] = E[\frac{1}{n} \sum_{i=1}^{n} X_i] = \mu$. Note that the only assumption needed is that the random variables have the same finite mean.

**Example 2e.** Expectation of a binomial random variable. Note that a binomial random variable is sum of $n$ Bernoulli trials. The expectation of a Bernoulli trial is $p$. So the expectation of sum of $n$ Bernoulli trials is $np$.

**Example 2q.** Let $X$ be a nonnegative, integer-valued random variable. Prove that $E[X] = \sum_{i=1}^{n} P[X \geq i]$. 

Read Example 2g, 2i, 2j
7.3 Covariance, Variance of Sums, and Correlations

Proposition 3.1
If $X$ and $Y$ are independent variables, then for any functions $h$ and $g$,

$$\text{E}[g(X)h(Y)] = \text{E}[g(X)]\text{E}[h(Y)].$$

Definition
The covariance between $X$ and $Y$, denoted by $\text{Cov}(X,Y)$, is defined by

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

One identity that is often helpful is that $\text{Cov}(X,Y) = \text{E}[XY] - \text{E}[X]\text{E}[Y]$.

Proposition 3.2
(i) $\text{cov}(X,Y) = \text{cov}(Y,X)$.
(ii) $\text{cov}(X,X) = \text{var}(X)$.
(iii) $\text{cov}(aX,Y) = a \text{cov}(Y,X)$.
(iv) $\text{cov}(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} \text{cov}(X_i,Y_j)$.

Example 3a. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables having expected value $\mu$ and variance $\sigma^2$. Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, and let

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2.$$

Compute $\text{var}(\bar{X}_n)$ and $\text{E}[S^2]$.

Definition
The correlation of two random variables $X$ and $Y$, denoted by $\rho(X,Y)$, is defined, as long as $0 < \text{var}(X) < \infty$ and $0 < \text{var}(Y) < \infty$, by $\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$.

Cauchy-Schwartz inequality generalizes to expectations:

$$\text{E}[XY] \leq \sqrt{\text{E}[X^2]\text{E}[Y^2]}.$$