Properties of Random Variables

1. Definitions

A **discrete random variable** is defined by a **probability distribution** that lists each possible outcome and the probability of obtaining that outcome. If the random variable x is discrete with n possible outcomes, here is its probablilty distribution.

Outcome	Probability
x_1	p_1
x_2	p_2
$\vdots$	$\vdots$
x_n	p_n

The commonly used discrete Binomial random variable shows the number of successes for n independent random trials when the probability of a single success is p. Here is the Binomial probability distribution f for obtaining k successes out of n trials:

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

Example: What is the probability distribution for the number of successes for shooting 3 basketball free throws when you are a 80% free throw shooter? Answer:

Outcome	Probability
0	0.008
1	0.096
2	0.384
3	0.512

A **continuous random variable** is defined by its **probability density function**. To find the probability that $a \leq x \leq b$ for the continuous random variable x, find the area under its probability density p:

$$P(a \leq x \leq b) = \int_a^b p(x)\, dx.$$

Note that for any value a, $P(x = a) = 0$ for a continuous random variable x.

The most popular continuous distribution is the normal distribution, which is discussed in Section 4.

2. Expected Value

The **expected value** of a random variable is the weighted average of its possible values. Each value is weighted by the probability that the outcome occurs.

If a discrete random variable x has outcomes $x_1, x_2, \cdots, x_n$, with probabilities $p_1, p_2, \cdots, p_n$, respectively, the expected value of x is

$$E(x) = \sum_{i=1}^{n} x_i p_i.$$

If a continuous random variable has the probability density p, its expected value is defined by

$$E(x) = \int_{-\infty}^{\infty} x p(x)\, dx.$$

In the justification of the properties of random variables later in this section, we assume continuous random variables. The justifications for discrete random variables are obtained by replacing the integrals with summations. For the remainder of this section, the letters x and y represent random variables and the letter c represents a constant.

Property 1: $\boxed{\mathrm{E}(x+y) = \mathrm{E}(x) + \mathrm{E}(y)}$

Justification: Let $p(x)$ and $q(y)$ be the probability density functions of the random variables x and y, respectively. Also let $r(x, y)$ be the joint density of x and y considered together. We have

$$p(x) = \int_{-\infty}^{\infty} r(x, y)\, dy \quad \text{and} \quad q(y) = \int_{-\infty}^{\infty} r(x, y)\, dx$$

$p(x)$ and $q(x)$ are called the **marginal densities** of $r(x, y)$.

Then

$$
\begin{aligned}
\mathrm{E}(x+y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)r(x,y)\,dy\,dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,r(x,y)\,dy\,dx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\,r(x,y)dy\,dx \\
&= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} r(x,y)\,dy\,dx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} r(x,y)dx\,dy \\
&= \int_{-\infty}^{\infty} x\,p(x)\,dx + \int_{-\infty}^{\infty} y\,q(x)\,dy \\
&= \mathrm{E}(x) + \mathrm{E}(y)
\end{aligned}
$$

$\blacksquare$

Property 2: $\boxed{\mathrm{E}(c) = c}$

Justification: c can be thought of as a discrete random variable that takes on the value c with probability 1. $\mathrm{E}(c)$ is then simply the value of c times 1. $\blacksquare$

Property 3: $\boxed{\mathrm{E}(cx) = c\,\mathrm{E}(x)}$

Justification: Let $p(x)$ be the probability density function of the random variable x. Then

$$\mathrm{E}(cx) = \int_{-\infty}^{\infty} cx\,p(x)\,dx = c\int_{-\infty}^{\infty} x\,p(x)\,dx = c\,\mathrm{E}(x)$$

$\blacksquare$

2. Independence

Two random variables x and y are **independent** if their joint density $r(x,y)$ factors into two one-dimensional densities: $r(x,y) = p(x)q(y)$.

Property 4: $\boxed{\text{If } x \text{ and } y \text{ are independent, } \mathrm{E}(xy) = \mathrm{E}(x)\,\mathrm{E}(y).}$

Justification:

$$
\begin{aligned}
\mathrm{E}(xy) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, r(x,y)\, dy\, dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, p(x)q(y)\, dy\, dx \\
&= \int_{-\infty}^{\infty} x\, p(x) \int_{-\infty}^{\infty} y\, q(y)\, dy\, dx \\
&= \int_{-\infty}^{\infty} x\, p(x)\, dx \int_{-\infty}^{\infty} y\, q(y)\, dy \\
&= \mathrm{E}(x)\,\mathrm{E}(y)
\end{aligned}
$$

∎

3. Variance

The **variance** of a random variable x is defined as $\mathrm{Var}(x) = \mathrm{E}(x - \mathrm{E}(x))^2$.

Property 5: $\boxed{\mathrm{Var}(x) = \mathrm{E}(x^2) - E(x)^2}$

Justification:

$$\mathrm{Var}(x) = \mathrm{E}(x - \mathrm{E}(x))^2 = \mathrm{E}(x^2) - 2\,\mathrm{E}(x)\,\mathrm{E}(x) + \mathrm{E}(x)^2 = \mathrm{E}(x^2) - \mathrm{E}(x)^2$$

∎

Property 6: $\boxed{\text{If } x \text{ and } y \text{ are independent, } \mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(y).}$

Justification:

$$
\begin{aligned}
\mathrm{Var}(x+y) &= \mathrm{E}((x+y)^2) - \mathrm{E}(x+y)^2 \\
&= \mathrm{E}(x^2 + 2xy + y^2) - (\mathrm{E}(x) + \mathrm{E}(y))^2 \\
&= \mathrm{E}(x^2) + 2\,\mathrm{E}(xy) + \mathrm{E}(y^2) - (\mathrm{E}(x)^2 + 2\,\mathrm{E}(x)\,\mathrm{E}(y) + \mathrm{E}(y)^2) \\
&= \mathrm{E}(x^2) + 2\,\mathrm{E}(x)\,\mathrm{E}(y) + \mathrm{E}(y^2) - (\mathrm{E}(x)^2 + 2\,\mathrm{E}(x)\,\mathrm{E}(y) + \mathrm{E}(y)^2) \\
&= \mathrm{E}(x^2) - \mathrm{E}(x)^2 + 2\,\mathrm{E}(x)\,\mathrm{E}(y) - 2\,\mathrm{E}(x)\,\mathrm{E}(y) + \mathrm{E}(y^2) - \mathrm{E}(y)^2 \\
&= \mathrm{Var}(x) + \mathrm{Var}(y)
\end{aligned}
$$

∎

Property 7: $\boxed{\mathrm{Var}(cx) = c^2\,\mathrm{Var}(x)}$

Justification:

$$\begin{aligned}\mathrm{Var}(cx) &= \mathrm{E}((cx)^2) - \mathrm{E}(cx)^2 = \mathrm{E}(c^2x^2) - (c\,\mathrm{E}(x))^2 \\ &= c^2\,\mathrm{E}(x^2) - c^2\,\mathrm{E}(x)^2 = c^2(\mathrm{E}(x^2) - \mathrm{E}(x)^2) \\ &= c^2\,\mathrm{Var}(x)\end{aligned}$$

∎

4. The Normal Density

The **normal distribution** with mean μ and variance σ^2 is defined by its density

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{2\sigma}\right)^2} \tag{1}$$

The expression

$$x \sim N(\mu, \sigma^2) \tag{2}$$

means that the random variable x has a normal distribution with mean μ and variance σ^2.

Ask the instructor for references if you are interested in a justification of (1) implies (2). A modern justification that $\mathrm{E}(x) = \mu$ and $\mathrm{Var}(x) = \sigma^2$ for a normal random variable x uses the moment generating function, which is defined as $M_x(t) = \mathrm{E}(e^{-tx})$. For a $N(\mu, \sigma^2)$ random variable, $M_x(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$.

5. The Unbiased Property of the Sample Mean

A statistic t is **unbiased** for a parameter θ if $\mathrm{E}(t) = \theta$.

Property 8: $\boxed{x \sim N(\mu, \sigma^2) \text{ implies } \bar{x} \text{ is unbiased for } \mu.}$

Justification:

Suppose that $x_i \sim N(\mu, \sigma^2)$ for each observation x_i in the random sample $x_1, \ldots, x_n$. Then $\mathrm{E}(x_i) = \mu$ and $\mathrm{Var}(x_i) = \sigma^2$. Furthermore,

$$\mathrm{E}(\bar{x}) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(x_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{n}{n}\mu = \mu$$

$\blacksquare$

Property 9: If the observations in the random sample x_1, x_2, $\cdots$, x_n are independent and $x_i \sim N(\mu, \sigma^2)$ for each observation x_i,

$$\boxed{\operatorname{Var}(\bar{x}) = \frac{\sigma^2}{n}}$$

Justification:

$$\begin{aligned}
\operatorname{Var}(\bar{x}) &= \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}(x_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n}{n^2}\sigma^2 = \frac{\sigma^2}{n}
\end{aligned}$$

$\blacksquare$

7. The Unbiased Property of the Sample Variance

The **sample standard deviation** s is defined as

$$s = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

The $n-1$ in the denominator is chosen so that s_x^2 will be unbiased for $\sigma_x^2 2$, as we see in this property:

Property 10: $\boxed{s_x^2 \text{ is unbiased for } \sigma_x^2.}$

Justification: Assume that we have a random sample $x_1, \ldots, x_n$, where all observations are independent, $\mathrm{E}(x_i) = \mu$ and $\operatorname{Var}(x_i) = \sigma^2$. for all observations. Since $\operatorname{Var}(x_i) = \mathrm{E}(x_i^2) - \mathrm{E}(x_i)^2$ by Property 4, $\sigma^2 = \mathrm{E}(x_i)^2 - \mu^2$, so $\mathrm{E}(x_i^2) = \sigma^2 + \mu^2$. Also, if $x_i \neq x_j$, using the fact that x_i and x_j are independent, by Property 3, $\mathrm{E}(x_i x_j) = \mathrm{E}(x_i)\,\mathrm{E}(x_j) = \mu\mu = \mu^2$. We then have

$$
\begin{aligned}
\mathrm{E}(\mathrm{SSE}) &= \mathrm{E}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{n}\left(x_i-\frac{1}{n}\sum_{j=1}^{n}x_j\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{n}\left(x_i^2-\frac{2}{n}\sum_{j=1}^{n}x_ix_j+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}x_jx_k\right)\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{n}x_i^2-\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_j+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}x_jx_k\right) \\
&= \sum_{i=1}^{n}\mathrm{E}(x_i^2)-\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}(x_ix_j)+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathrm{E}(x_jx_k) \\
&= \sum_{i=1}^{n}\mathrm{E}(x_i^2)-\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}(x_ix_j)+\frac{n}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathrm{E}(x_jx_k) \\
&= \sum_{i=1}^{n}\mathrm{E}(x_i^2)-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}(x_ix_j) \\
&= n(\sigma^2+\mu^2)-\frac{1}{n}[n(\sigma^2+\mu^2)+n(n-1)\mu^2] \\
&= n\sigma^2+n\mu^2-\sigma^2+\mu^2+(n-1)\mu^2] \\
&= (n-1)\sigma^2
\end{aligned}
$$

Therefore,

$$
\mathrm{E}(s_x)=\mathrm{E}\left(\frac{\mathrm{SSE}}{n-1}\right)=\frac{1}{n-1}\,\mathrm{E}\,(\mathrm{SSE})=\frac{1}{n-1}(n-1)\sigma^2=\sigma^2.
$$

∎

Random Vectors

Property 11: $\boxed{\mathrm{E}(\mathbf{v}+\mathbf{w})=\mathrm{E}(\mathbf{v})+\mathrm{E}(\mathbf{w}).}$

Justification:

$$
\begin{aligned}
\mathrm{E}(\mathbf{v}+\mathbf{w}) &= \mathrm{E}\begin{pmatrix} v_1 + w_1 \\ \vdots \\ v_n + w_n \end{pmatrix} = \begin{pmatrix} \mathrm{E}(v_1 + w_1) \\ \vdots \\ \mathrm{E}(v_n + w_n) \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{E}(v_1) + \mathrm{E}(w_1) \\ \vdots \\ \mathrm{E}(v_n) + \mathrm{E}(w_n) \end{pmatrix} = \begin{pmatrix} \mathrm{E}(v_1) \\ \vdots \\ \mathrm{E}(v_n) \end{pmatrix} + \begin{pmatrix} \mathrm{E}(w_1) \\ \vdots \\ \mathrm{E}(w_n) \end{pmatrix} \\
&= \mathrm{E}(\mathbf{v}) + \mathrm{E}(\mathbf{w})
\end{aligned}
$$

$\blacksquare$

Property 12: If $\mathbf{b}$ is a constant vector, $\boxed{\mathrm{E}(\mathbf{b}) = \mathbf{b}.}$

Justification:

$$
\mathrm{E}(\mathbf{b}) = \mathrm{E}\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} \mathrm{E}(b_1) \\ \vdots \\ \mathrm{E}(b_n) \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \mathbf{b}
$$

$\blacksquare$

Property 13: $\boxed{\mathrm{E}(\mathbf{A}\mathbf{v}) = \mathbf{A}\,\mathrm{E}(\mathbf{w}).}$

Justification: Let $\mathbf{A}$ be an $m \times n$ matrix and $\mathbf{v}$ be an $n \times 1$ vector.

$$
\begin{aligned}
\mathrm{E}(\mathbf{A}\mathbf{v}) &= \mathrm{E}\left(\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}\right) \\
&= \mathrm{E}\begin{pmatrix} \sum_{i=1}^{n} a_{1i}v_i \\ \vdots \\ \sum_{i=1}^{n} a_{mi}v_i \end{pmatrix} = \begin{pmatrix} \mathrm{E}\left(\sum_{i=1}^{n} a_{1i}v_i\right) \\ \vdots \\ \mathrm{E}\left(\sum_{i=1}^{n} a_{mi}v_i\right) \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^{n} a_{1i}\,\mathrm{E}(v_i) \\ \vdots \\ \sum_{i=1}^{n} a_{mi}\,\mathrm{E}(v_i) \end{pmatrix} = \begin{pmatrix} a_{1i} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} \mathrm{E}(v_1) \\ \vdots \\ \mathrm{E}(v_n) \end{pmatrix} \\
&= \mathbf{A}\,\mathrm{E}(\mathbf{v})
\end{aligned}
$$

∎

Property 14: $\boxed{\mathrm{Cov}(\mathbf{v}+\mathbf{b}) = \mathrm{Cov}(\mathbf{v}).}$

Justification:

$$
\begin{aligned}
\mathrm{Cov}(\mathbf{v}+\mathbf{b}) &= \begin{pmatrix} \mathrm{Var}(v_1+b) & \cdots & \mathrm{Cov}(v_n+b, v_1+b) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(v_1+b, v_n+b) & \cdots & \mathrm{Var}(v_n+b) \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{Var}(v_1) & \cdots & \mathrm{Cov}(v_n, v_1) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(v_1, v_n) & \cdots & \mathrm{Var}(v_n) \end{pmatrix} \\
&= \mathrm{Cov}(\mathbf{v})
\end{aligned}
$$

∎

Property 15: $\boxed{\mathrm{Cov}(\mathbf{A}\mathbf{v}) = \mathbf{A}\,\mathrm{Cov}(\mathbf{v})\mathbf{A}^T.}$

Justification: Let $\mathrm{Cov}(v_i, v_j) = \sigma_{ij}$. Then

$$
\begin{aligned}
\mathrm{Cov}(\mathbf{A}\mathbf{v}) &= \mathrm{Cov}\left(\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}\right) \\
&= \mathrm{Cov}\begin{pmatrix} \sum_{i=1}^n a_{1i}v_i \\ \vdots \\ \sum_{i=1}^n a_{ni}v_i \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{Var}\left(\displaystyle\sum_{i=1}^n a_{1i}v_i\right) & \cdots & \mathrm{Cov}\left(\displaystyle\sum_{i=1}^n a_{1i}v_i, \sum_{i=1}^n a_{ni}v_i,\right) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}\left(\displaystyle\sum_{i=1}^n a_{1i}v_i, \sum_{i=1}^n a_{ni}v_i,\right) & \cdots & \mathrm{Var}\left(\displaystyle\sum_{i=1}^n a_{nn}v_i\right) \end{pmatrix} \\
&= \begin{pmatrix} \displaystyle\sum_{i=1}^n\sum_{j=1}^n a_{1i}a_{1j}\sigma_{ij} & \cdots & \displaystyle\sum_{i=1}^n\sum_{j=1}^n a_{1i}a_{nj}\sigma_{ij} \\ \vdots & \ddots & \vdots \\ \displaystyle\sum_{i=1}^n\sum_{j=1}^n a_{ni}a_{1j}\sigma_{ij} & \cdots & \displaystyle\sum_{i=1}^n\sum_{j=1}^n a_{ni}a_{nj}\sigma_{ij} \end{pmatrix} \\
&= \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} \displaystyle\sum_{i=1}^n a_{1j}\sigma_{1j} & \cdots & \displaystyle\sum_{i=1}^n a_{nj}\sigma_{1j} \\ \vdots & \ddots & \vdots \\ \displaystyle\sum_{i=1}^n a_{1j}\sigma_{nj} & \cdots & \displaystyle\sum_{i=1}^n a_{nj}\sigma_{nj} \end{pmatrix} \\
&= \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \dots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}\begin{pmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \dots & \vdots \\ a_{1n} & \cdots & a_{nn} \end{pmatrix} \\
&= \mathbf{A}\,\mathrm{Cov}(\mathbf{v})\mathbf{A}^T
\end{aligned}
$$

∎