Probability Distributions

If $X$ is a random variable on sample space $S$, then the probability that $X$ takes on the value $c$ is

$$Pr(X = c) = Pr\{s \in S \mid X(s) = c\}$$

Similarly,

$$Pr(X \leq c) = Pr\{s \in S \mid X(s) \leq c\}.$$

This makes sense since the range of $X$ is the real numbers.

**Example:** In the coin example,

$$Pr(\# H = 2) = \frac{4}{9} \text{ and } Pr(\# H \leq 1) = \frac{5}{9}$$

Given a probability measure $Pr$ on a sample space $S$ and a random variable $X$, the probability distribution associated with $X$ is $f_X(x) = Pr(X = x)$.

- $f_X$ is a probability measure on the real numbers.

The cumulative distribution associated with $X$ is $F_X(x) = Pr(X \leq x)$.

The Finite Uniform Distribution

The finite uniform distribution is an equiprobable distribution. If $S = \{x_1, \ldots, x_n\}$, where $x_1 < x_2 < \ldots < x_n$, then:

$$f(x_k) = \frac{1}{n}$$

$$F(x_k) = \frac{k}{n}$$

An Example With Dice

Suppose $S$ is the sample space corresponding to tossing a pair of fair dice: $\{(i, j) \mid 1 \leq i, j \leq 6\}$.

Let $X$ be the random variable that gives the sum:

- $X(i, j) = i + j$

Given a probability measure $Pr$ on a sample space $S$ and a random variable $X$, the probability distribution associated with $X$ is $f_X(x) = Pr(X = x)$.

- $f_X$ is a probability measure on the real numbers.

The cumulative distribution associated with $X$ is $F_X(x) = Pr(X \leq x)$.

The Binomial Distribution

Suppose there is an experiment with probability $p$ of success and thus probability $q = 1 - p$ of failure.

- For example, consider tossing a biased coin, where $Pr(h) = p$. Getting “heads” is success, and getting tails is failure.

Suppose the experiment is repeated independently $n$ times.

- For example, the coin is tossed $n$ times.

This is called a sequence of Bernoulli trials.

Key features:

- Only two possibilities: success or failure.
- Probability of success does not change from trial to trial.
- The trials are independent.
What is the probability of $k$ successes in $n$ trials?

Suppose $n = 5$ and $k = 3$. How many sequences of 5 coin tosses have exactly three heads?

- $hhhtt$
- $hhtht$
- $hhtth$

$C(5, 3)$ such sequences!

What is the probability of each one?

$$p^3(1 - p)^2$$

Therefore, probability is $C(5, 3)p^3(1 - p)^2$.

Let $B_{n,p}(k)$ be the probability of getting $k$ successes in $n$ Bernoulli trials with probability $p$ of success.

$$B_{n,p}(k) = C(n, k)p^k(1 - p)^{n-k}$$

Not surprisingly, $B_{n,p}$ is called the Binomial Distribution.

### The Poisson Distribution

A large call center receives, on average, $\lambda$ calls/minute.

- What is the probability that exactly $k$ calls come during a given minute?

Understanding this probability is critical for staffing!

- Similar issues arise if a printer receives, on average $\lambda$ jobs/minute, a site gets $\lambda$ hits/minute, ... 

This is modelled well by the Poisson distribution with parameter $\lambda$:

$$f_\lambda(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

- $f_\lambda(0) = e^{-\lambda}$
- $f_\lambda(1) = e^{-\lambda} \lambda$
- $f_\lambda(2) = e^{-\lambda} \frac{\lambda^2}{2}$

$e^{-\lambda}$ is a normalization constant, since

$$1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{3!} + \cdots = e^\lambda$$

### Deriving the Poisson

Poisson distribution $= \lim$ of binomial distributions.

Suppose at most one call arrives in each second.

- Since $\lambda$ calls come each minute, expect about $\lambda/60$ each second.
- The probability that $k$ calls come is $B_{60,\lambda/60}(k)$

This model doesn’t allow more than one call/second.

What’s so special about 60? Suppose we divide one minute into $n$ time segments.

- Probability of getting a call in each segment is $\lambda/n$.
- Probability of getting $k$ calls in a minute is $B_{n,\lambda/n}(k)$

Now let $n \to \infty$:

- $\lim_{n \to \infty} (1 - \frac{\lambda}{n})^n = e^{-\lambda}$
- $\lim_{n \to \infty} \left( \frac{n!}{(n-k)!} \right)^k \frac{(1 - \frac{\lambda}{n})^n}{(1 - \frac{\lambda}{n-k})^k} = 1$

Conclusion: $\lim_{n \to \infty} B_{n,\lambda/n}(k) = e^{-\lambda} \frac{\lambda^k}{k!}$

### New Distributions from Old

If $X$ and $Y$ are random variables on a sample space $S$, so is $X + Y$, $X + 2Y$, $XY$, $\sin(X)$, etc.

For example,

- $(X + Y)(s) = X(s) + Y(s)$.
- $\sin(X)(s) = \sin(X(s))$

Note $\sin(X)$ is a random variable: a function from the sample space to the reals.
Some Examples

Example 1: A fair die is rolled. Let $X$ denote the number that shows up. What is the probability distribution of $Y = X^2$?

$$
\{s : Y(s) = k\} = \{s : X^2(s) = k\} = \{s : X(s) = \sqrt{k}\} \cup \{s : X(s) = -\sqrt{k}\}.
$$

Conclusion: $f_Y(k) = f_X(\sqrt{k}) + f_X(-\sqrt{k})$.

So $f_Y(1) = f_X(4) = f_X(9) = \cdots f_X(36) = 1/6$. $f_Y(k) = 0$ if $k \not\in \{1, 4, 9, 16, 25, 36\}$.

Example 2: A coin is flipped. Let $X$ be 1 if the coin shows $H$ and -1 if $T$. Let $Y = X^2$.

- In this case $Y \equiv 1$, so $\Pr(Y = 1) = 1$.

Example 3: If two dice are rolled, let $X$ be the number that comes up on the first dice, and $Y$ the number that comes up on the second.

- Formally, $X((i, j)) = i$, $Y((i, j)) = j$.

The random variable $X + Y$ is the total number showing.

Independent random variables

In a roll of two dice, let $X$ and $Y$ record the numbers on the first and second die respectively.

- What can you say about the events $X = 3$, $Y = 2$?
- What about $X = i$ and $Y = j$?

Definition: The random variables $X$ and $Y$ are independent if for every $x$ and $y$ the events $X = x$ and $Y = y$ are independent.

Example: $X$ and $Y$ above are independent.

Definition: The random variables $X_1, X_2, \ldots, X_n$ are mutually independent if, for every $x_1, x_2, \ldots, x_n$

$$
\Pr(X_1 = x_1 \cap \ldots \cap X_n = x_n) = \Pr(X_1 = x_1) \ldots \Pr(X_n = x_n)
$$

Example: $X_k$, the success indicators in $n$ Bernoulli trials, are independent.

Pairwise vs. mutual independence

Mutual independence implies pairwise independence; the converse may not be true.

Example 1: A ball is randomly drawn from an urn containing 4 balls: one blue, one red, one green and one multicolored (red + blue + green)

- Let $X_1, X_2$ and $X_3$ denote the indicators of the events the ball has (some) blue, red and green respectively.

- $\Pr(X_i = 1) = 1/2$, for $i = 1, 2, 3$

$X_1$ and $X_2$ independent:

<table>
<thead>
<tr>
<th>$X_1 = 0$</th>
<th>$X_1 = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_2 = 0$</td>
<td>1/4</td>
</tr>
<tr>
<td>$X_2 = 1$</td>
<td>1/4</td>
</tr>
</tbody>
</table>

Similarly, $X_1$ and $X_3$ are independent; so are $X_2$ and $X_3$. Are $X_1$, $X_2$ and $X_3$ independent? No!

$$
\Pr(X_1 = 1) \cap X_2 = 1 \cap X_3 = 1) = 1/4
\Pr(X_1 = 1) \Pr(X_2 = 1) \Pr(X_3 = 1) = 1/8.
$$

Example 2: Suppose $X_1$ and $X_2$ are bits (0 or 1) chosen uniformly at random; $X_3 = X_1 \oplus X_2$.

- $X_1, X_2$ are independent, as are $X_1, X_3$ and $X_2, X_3$
- But $X_1, X_2$ and $X_3$ are not mutually independent

$X_1$ and $X_2$ together determine $X_3$!
The distribution of $X + Y$

Suppose $X$ and $Y$ are independent random variables whose range is included in $\{0, 1, \ldots, n\}$. For $k \in \{0, 1, \ldots, 2n\}$,

$$(X + Y = k) = \bigcup_{j=0}^{k} ((X = j) \cap (Y = k - j)) .$$

Note that some of the events might be empty

• E.g., $X = k$ is bound to be empty if $k > n$.

This is a disjoint union so

$$
\Pr(X + Y = k) = \sum_{j=0}^{k} \Pr(X = j) \Pr(Y = k - j) \quad \text{[by independence]}
$$

Example: The Sum of Binomials

Suppose $X$ has distribution $B_{n,p}$, $Y$ has distribution $B_{m,p}$, and $X$ and $Y$ are independent.

$$
\Pr(X + Y = k) = \sum_{j=0}^{k} \Pr(X = j) \Pr(Y = k - j) \quad \text{[sum rule]}
$$

$$
= \sum_{j=0}^{k} \Pr(X = j) \Pr(Y = k - j) \quad \text{[independence]}
$$

$$
= \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j} \binom{m}{k-j} p^{k-j} (1-p)^{m-k+j}
$$

$$
= \binom{n+m}{k} p^k (1-p)^{n+m-k}
$$

Thus, $X + Y$ has distribution $B_{n+m,p}$.

An easier argument: Perform $n + m$ Bernoulli trials. Let $X$ be the number of successes in the first $n$ and let $Y$ be the number of successes in the last $m$. $X$ has distribution $B_{n,p}$, $Y$ has distribution $B_{m,p}$. $X$ and $Y$ are independent, and $X + Y$ is the number of successes in all $n + m$ trials, and so has distribution $B_{n+m,p}$.

Expected Value

Suppose we toss a biased coin, with $\Pr(h) = 2/3$. If the coin lands heads, you get $1$; if the coin lands tails, you get $3$. What are your expected winnings?

• $2/3$ of the time you get $1$;
  $1/3$ of the time you get $3$

• $(2/3 \times 1) + (1/3 \times 3) = 5/3$

What’s a good way to think about this? We have a random variable $W$ (for winnings):

• $W(h) = 1$
• $W(t) = 3$

The expectation of $W$ is

$$
E(W) = \Pr(h)W(h) + \Pr(t)W(t)
$$

$$
= \Pr(W = 1) \times 1 + \Pr(W = 3) \times 3
$$

More generally, the expected value of random variable $X$ on sample space $S$ is

$$
E(X) = \sum_{x \in S} x \Pr(X = x)
$$

An equivalent definition:

$$
E(X) = \sum_{s \in S} X(s) \Pr(s)
$$
What is $E(B_{n,p})$, the expectation for the binomial distribution $B_{n,p}$?

- How many heads do you expect to get after $n$ tosses of a biased coin with $\Pr(h) = p$?

**Method 1:** Use the definition and crank it out:

\[
E(B_{n,p}) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}
\]

This looks awful, but it can be calculated ...

**Method 2:** Use Induction; break it up into what happens on the first toss and on the later tosses.

- On the first toss you get heads with probability $p$ and tails with probability $1-p$. On the last $n-1$ tosses, you expect $E(B_{n-1,p})$ heads. Thus, the expected number of heads is:

\[
E(B_{n,p}) = p(1 + E(B_{n-1,p})) + (1-p)(E(B_{n-1,p}))
\]

\[
E(B_{n,p}) = p + E(B_{n-1,p})
\]

Now an easy induction shows that $E(B_{n,p}) = np$.

There's an even easier way . . .

**Example 1:** Back to the expected value of tossing two dice:

Let $X_1$ be the count on the first die, $X_2$ the count on the second die, and let $X$ be the total count.

Notice that

\[
E(X_1) = E(X_2) = \frac{(1 + 2 + 3 + 4 + 5 + 6)}{6} = 3.5
\]

\[
E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7
\]

**Example 2:** Back to the expected value of $B_{n,p}$.

Let $X$ be the total number of successes and let $X_k$ be the outcome of the $k$th experiment, $k = 1, \ldots, n$:

\[
E(X_k) = p \cdot 1 + (1-p) \cdot 0 = p
\]

\[
X = X_1 + \cdots + X_n
\]

Therefore

\[
E(X) = E(X_1) + \cdots + E(X_n) = np
\]

**Expectation of Poisson Distribution**

Let $X$ be Poisson with parameter $\lambda$: $f_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ for $k \in \mathbb{N}$.

\[
E(X) = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!}
\]

\[
= \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!}
\]

\[
= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}
\]

\[
= \lambda e^{-\lambda} e^{\lambda} = \lambda
\]

Does this make sense?

- Recall that, for example, $X$ models the number of incoming calls for a tech support center whose average rate per minute is $\lambda$. 

**Expectation is Linear**

**Theorem:** $E(X + Y) = E(X) + E(Y)$

**Proof:** Recall that

\[
E(X) = \sum_{s \in S} \Pr(s) X(s)
\]

Thus,

\[
E(X + Y) = \sum_{s \in S} \Pr(s)(X(s) + Y(s))
\]

\[
= \sum_{s \in S} \Pr(s)X(s) + \sum_{s \in S} \Pr(s)Y(s)
\]

\[
= E(X) + E(Y).
\]
Geometric Distribution

Consider a sequence of Bernoulli trials. Let $X$ denote the number of the first successful trial.

- E.g., the first time you see heads $X$ has a geometric distribution.

\[ f_X(k) = (1 - p)^{k-1} p \quad k \in \mathbb{N}^+. \]

- The probability of seeing heads for the first time on the $k$th toss is the probability of getting $k - 1$ tails followed by heads.

- This is also called a negative binomial distribution of order 1.
  - The negative binomial of order $n$ gives the probability that it will take $k$ trials to have $n$ successes.