Probability Distributions

- There is a natural probabilistic structure induced on a random variable X defined on Ω:
  - · The set $\{\omega \in \Omega : X(\omega) = c\}$ is an event.
  - · So we can ask for

$$\Pr(X = c) = \Pr(\{\omega \in \Omega : X(\omega) = c\}).$$

  Example. A biased coin ($\Pr(H) = 2/3$) is flipped twice.

  - · Let X count the number of heads:

  $\Pr(X = 0) = \Pr(\{TT\}) = (1/3)^2 = 1/9.$
  $\Pr(X = 1) = \Pr(\{HT, TH\}) = 2 \cdot 1/3 \cdot 2/3 = 4/9.$
  $\Pr(X = 2) = \Pr(\{HH\}) = (2/3)^2 = 4/9.$

- Similarly we might be interested in:

$$\Pr(X \leq c) = \Pr(\{\omega \in \Omega : X(\omega) \leq c\},$$

  and more generally, for any $T \subset \mathbb{R}$:

$$\Pr(X \in T) = \Pr(\{\omega \in \Omega : X(\omega) \in T\}.$$

- In our coin example,

  $\Pr(X \leq 1) = \Pr(\{TT, HT, TH\}) = 1/9+4/9 = 5/9.$

- **Def.** The function $f_X(x) := \Pr(X = x)$ is called the *probability mass function* (pmf), or the *probability distribution*, or the *density function* of X.
- In our coin example,

$$f_X(x) = \begin{cases} 1/9 & x = 0 \\ 4/9 & x = 1 \\ 4/9 & x = 2 \\ 0 & \text{otherwise} \end{cases}.$$

- The pmf conveys all the probabilistic information that is relevant to X.
- A similar function that can yield the same information is:
- **Def.** The *cummulative distribution function* (cdf) of X is $F_X(x) := \Pr(X \leq x)$.
- In our coin example,

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/9 & x \in [0, 1) \\ 5/9 & x \in [1, 2) \\ 1 & x \geq 2 \end{cases}.$$

- Graphically:

An Example With Dice

- A pair of fair dice is rolled.
- Let X be the random variable that gives the sum of the faces.
- Find the pmf of X.
  - $\Omega = \{(i, j) | \, 1 \leq i, j \leq 6\}$ with $\Pr\{(i, j)\} = 1/36$.
  - $X(i, j) = i + j$.

$$
\begin{aligned}
f_X(2) &= \Pr(X = 2) = \Pr(\{(1,1)\}) = 1/36 \\
f_X(3) &= \Pr(X = 3) = \Pr(\{(1,2),(2,1)\}) = 2/36 \\
&\vdots \\
f_X(7) &= \Pr(X = 7) = \Pr(\{(1,6),\ldots,(6,1)\}) = 6/36 \\
&\vdots \\
f_X(12) &= \Pr(X = 12) = \Pr(\{(6,6)\}) = 1/36
\end{aligned}
$$

- The cdf can readily be found from the pmf. For example, for $x \in [3, 4)$: $F_X(x) = f_X(2) + f_X(3) = 3/36$.

cdf from pmf and vice versa

- The pmf of X can be derived from its cdf:
  - For simplicity assume that X cannot attain any value between $y < x$.
  - $\{X \leq x\} = \{X \leq y\} \cup \{X = x\}$
  - $\Rightarrow \Pr\{X \leq x\} = \Pr\{X \leq y\} + \Pr\{X = x\}$
  - $\Rightarrow f_X(x) = F_X(x) - F_X(y)$
- Similarly we can derive the cdf from X's pmf:
  - $\{X \leq x\} = \cup_{y \leq x}\{X = y\}$
  - $\Rightarrow F_X(x) = \sum_{y \leq x} f_X(y)$

The Finite Uniform Distribution

- The finite uniform distribution is an equiprobable distribution.
- Suppose $X : \Omega \mapsto \{x_1, x_2, \ldots, x_n\}$ where $x_i < x_{i+1}$.
- Then, with $x_{n+1} := \infty$

$$f(x) = \begin{cases} 1/n & x = x_k \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 & x < x_1 \\ k/n & x \in [x_k, x_{k+1}) \end{cases}$$

- How do the graphs look like?
- Can you think of an example?
  - Fair die, fair coin.

The Binomial Distribution

- Suppose there is an experiment with probability p of success and thus probability $q = 1 - p$ of failure.
- **Examples.**
  - Tossing a coin ($\Pr(H) = p$): getting H is success, and getting T is failure.
  - Guessing the answer in a quiz of multiple choice questions (with four possible answers for each).
- Suppose the experiment is repeated independently n times.
  - The coin is tossed n times.
  - There are n questions.
- This is called a sequence of (n) *Bernoulli trials*.
- Key features:
  - Only two possibilities: success or failure.
  - Probability of success does not change from trial to trial.
  - The trials are independent.
- **Def.** A *binomial* random variable is one which counts the number of successes in n Bernoulli trials.

- Want to find $B_{n,p}(k)$, the pmf of a binomial random variable or the binomial distribution.
- Consider the coin example with $n = 5$ and $k = 3$.
  - $\{X = 3\} = \{HHHTT\} \cup \{HHTHT\} \ldots,$ where the union extends over all different sequences with exactly 3 Hs.
  - So,
    $\Pr(\{X = 3\}) = \Pr\{HHHTT\} + \Pr\{HHTHT\} + \ldots$
  - $\Pr(HHHTT) = p^3 q^2$:
    $$\{HHHTT\} = \{H - - - -\} \cap \{-H - - -\} \cap \{- - H - -\} \cap \{- - - T -\} \cap \{- - - - T\}$$
  - What is the probability of $HTHTH$?
  - So the probability of every sequence with exactly three Hs is $p^3 q^2$.
  - How many such sequences are there?
  - $\binom{5}{3}$
  - Therefore, $\Pr(\{X = 3\}) = B_{5,p}(3) = \binom{5}{3} p^3 q^2$.
- More generally, the probability of getting k successes in n Bernoulli trials with probability p of success is:

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

The Poisson Distribution

- The number of calls per minute to a tech support center between 5-8 pm is essentially constant (λ).
- What is the probability that on an ordinary day exactly k calls will come in between 6-6:01 pm?
- Binomial approximation:
  - If on average a coin lands H μ times in 60 flips then the probability that the next flip is H is $\mu/60$.
  - Divide the minute into 60 secs.
  - The probability that a call will arrive within any second is roughly $\lambda/60$.
  - The seconds are independent.
  - What is the probability that k out of the 60 seconds will be "successful"?
  - $B_{60,\lambda/60}(k)$.
  - What have we neglected?
  - There might be more than one call per second.
  - Divide the minute further into n equal intervals.
  - Approximate the probability by $B_{n,\lambda/n}(k)$.
  - What happens when $n \to \infty$?

$$
\begin{aligned}
B_{n,\lambda/n}(k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n!}{k!(n-k)!}\frac{1}{n^k}\lambda^k\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \left(1-\frac{\lambda}{n}\right)^n\frac{\lambda^k}{k!}\times \\
&\qquad \frac{n}{n}\frac{n-1}{n}\frac{n-2}{n}\cdots\frac{n-k+1}{n}\left(1-\frac{\lambda}{n}\right)^{-k} \\
&\xrightarrow[n\to\infty]{} e^{-\lambda}\frac{\lambda^k}{k!},
\end{aligned}
$$

where the latter is based on

$$
\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x.
$$

This is the Poisson distribution:

$$
f_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!} \qquad k = 0, 1, 2, \ldots
$$

New Distributions from Old

- Suppose X and Y are random variables on a sample space Ω.
- We can form new distributions from them: X^2, $\sin X$, $X + Y$, $X + 2Y$, XY, etc.
- For example,
  - $\cdot$ $Y = \sin X$ is defined by $Y(\omega) := \sin(X(\omega))$
  - $\cdot$ $Z = X + Y$ as $Z(\omega) := X(\omega) + Y(\omega)$.

Examples

- A fair die is rolled.
  - Let X denote the number that shows up.
  - What is the pmf of $Y = X^2$?

$$\begin{aligned}\{Y = k\} &= \{X^2 = k\} \\ &= \{X = -\sqrt{k}\} \cup \{X = \sqrt{k}\}.\end{aligned}$$

$$\begin{aligned}f_Y(k) &= f_X(\sqrt{k}) + f_X(-\sqrt{k}) \\ &= \begin{cases} 1/6 & k = i^2\,, i \in \{1, 2, \ldots, 6\} \\ 0 & \text{otherwise} \end{cases}.\end{aligned}$$

- A coin is flipped.
  - Let X be 1 if the coin shows H and -1 if T.
  - Let $Y = X^2$.
  - In this case $Y \equiv 1$.
- Two dice are rolled.
  - Let X be the number that comes up on the first die, and Y the number that comes up on the second.
  - Formally, $X((i, j)) = i$, $Y((i, j)) = j$.

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

- We toss a biased coin n times (more generally, we perform n Bernoulli trials).
  - X_k describes the outcome of the kth trial: $X_k = 1$ if it's heads (success), and 0 otherwise.
  - $\sum_{k=1}^{n} X_k$ describes the number of successes in n Bernoulli trials.

Independent random variables

- 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 $\{X = i\}$, $\{Y = j\}$?
- **Def.** 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.
- **Cor.** X and Y are independent if and only if

  $$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y) \qquad \forall x, y.$$

- **Def.** The random variables $X_1, X_2, \dots, X_n$ are independent if for every $x_1, x_2 \dots, x_n$

  $$\Pr(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \Pr(X_1 = x_1)\Pr(X_2 = x_2) \dots \Pr(X_n = x_n)$$

  Example. $\{X_k\}$, the success indicators in n Bernoulli trials are independent.

Pairwise independence does not imply independence

- A ball is randomly drawn from an urn containing 4 balls: blue, red, green and one which has all three colors.
- Let X_1, X_2 and X_3 denote the indicators of the events: the ball has blue, red and green respectively.
- What is $\Pr(X_1 = 1)$?
- $2/4 = 1/2$ and same for $\Pr(X_i = 1)$.
- Are X_1 and X_2 independent?
-

	$X_1 = 0$	$X_1 = 1$
$X_2 = 0$	$1/4$	$1/4$
$X_2 = 1$	$1/4$	$1/4$

- Same for X_1 and X_3, and for X_2 and X_3.
- Are X_1, X_2 and X_3 independent?
- No:

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

- The same warning applies for independent events: $\{X_i = 1\}$ are only pairwise independent.

Convolution, or the pmf of $X+Y$

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

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

- Note that many of the events might be empty.
- This is a disjoint union so

$$\begin{aligned}\Pr(X+Y=k) &= \sum_{j=0}^{k} \Pr(X=j, Y=k-j)\\ &= \sum_{j=0}^{k} \Pr(X=j)\Pr(Y=k-j).\end{aligned}$$

- In other words,

$$f_{X+Y}(k) = \sum_{j=0}^{k} f_X(j) f_Y(k-j).$$

- The right hand side is called the *convolution* of f_X and f_Y and is denoted by $f_X * f_Y$.

Example: sum of binomials

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

$$
\begin{aligned}
\Pr(X+Y=k) &= \sum_{j=0}^{k} \Pr(X=j)\Pr(Y=k-j) \\
&= \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j} \binom{m}{k-j} p^{k-j} (1-p)^{m-k+j} \\
&= \sum_{j=0}^{k} \binom{n}{j} \binom{m}{k-j} p^k (1-p)^{n+m-k} \\
&= p^k (1-p)^{n+m-k} \sum_{j=0}^{k} \binom{n}{j} \binom{m}{k-j} \\
&= \binom{n+m}{k} p^k (1-p)^{n+m-k}
\end{aligned}
$$

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.
- Let Y be the number of successes in the last m.
- On the one hand, $X + Y$ is the number of successes in all $n + m$ trials, and so has distribution $B_{n+m,p}$.
- On the other hand,
    - · X has distribution $B_{n,p}$
    - · Y has distribution $B_{m,p}$
    - · X and Y are independent (why?)
    - · So $X + Y$ has the distribution we are looking for.

Expectation, or 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
  - $(2/3 \times 1) + (1/3 \times 3) = 5/3$
- Formally, we have a random variable W (for winnings): $W(H) = 1$, $W(T) = 3$.
- The expectation of W is

$$\begin{aligned} E(W) &= \Pr(H)W(H) + \Pr(T)W(T) \\ &= \Pr(W = 1) \times 1 + \Pr(W = 3) \times 3 \end{aligned}$$

- **Def.** The *expectation* of random variable X is

$$E(X) = \sum_x x f_X(x).$$

- In other words, expectation is a weighted average.
- Technically we require that $\sum_x |x| f_X(x) < \infty$

Examples

- What is the expected count when two dice are tossed?
  - Let X be the count.

$$\begin{aligned} E(X) &= \sum_{i=2}^{12} i f_X(i) \\ &= 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + \cdots + 7\frac{6}{36} + \cdots + 12\frac{1}{36} \\ &= \frac{252}{36} \\ &= 7. \end{aligned}$$

- Let X be the indicator function of a success in one Bernoulli trial: $X = \begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$.
  -

$$E(X) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

  - Note that if $p \in (0, 1)$ X cannot attain its expected value.
- Since $E(X) = \sum_x x f_X(x)$ is defined in terms of the pmf of X one can talk about the expectation of a distribution.

Expectation of Binomials

- What is $E(B_{n,p})$, the expectation of the binomial distribution $B_{n,p}$?
- How many heads do you expect to get after n tosses of a biased coin with $\Pr(H) = p$?

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

- Note that

$$\begin{aligned} k\binom{n}{k} &= k\frac{n!}{k!(n-k)!} \\ &= n\frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} = n\binom{n-1}{k-1} \end{aligned}$$

- So

$$\begin{aligned} E(B_{n,p}) &= \sum_{k=1}^{n} n\binom{n-1}{k-1} p \cdot p^{k-1}(1-p)^{(n-1)-(k-1)} \\ &= np\sum_{j=0}^{n-1}\binom{n-1}{j} p^j (1-p)^{(n-1)-j} \\ &= np[p+(1-p)]^{n-1} \\ &= np. \end{aligned}$$

Expectation of Poisson distribution

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

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

- 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 λ.

Expectation of geometric distribution

- We observe a sequence of Bernoulli trials ($0 < p < 1$).
- Let X denote the number of the first succssful trial.
- X has a *geometric* distribution.

$$f_X(k) = (1-p)^{k-1} p \qquad k \in \mathbb{N}^+.$$

- What is the probability that X is finite?

$$\begin{aligned}\sum_{k=1}^{\infty} f_X(k) &= \sum_{k=1}^{\infty} (1-p)^{k-1} p \\ &= p \sum_{j=0}^{\infty} (1-p)^j \\ &= p \frac{1}{1-(1-p)} = 1.\end{aligned}$$

- What is $E(X)$? With $q = 1 - p$

$$\begin{aligned}E(X) &= \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} p = p \sum_{k=0}^{\infty} \frac{d}{dq} q^k \\ &= p \cdot \frac{d}{dq} \left(\sum_{k=0}^{\infty} q^k \right) = p \cdot \frac{d}{dq} \left(\frac{1}{1-q} \right) \\ &= p \cdot \frac{1}{(1-q)^2} = p \cdot \frac{1}{p^2} = \frac{1}{p}.\end{aligned}$$

The Expectation of $X + Y$

- **Claim.** $E(X) = \sum_{\omega \in \Omega} X(\omega) \Pr(\omega)$.
- **Proof.** Note that

$$
\begin{aligned}
f_X(x) &= \Pr(\{\omega \in \Omega : X(\omega) = x\}) \\
&= \sum_{\{\omega \in \Omega : X(\omega) = x\}} \Pr(\omega).
\end{aligned}
$$

$$
\begin{aligned}
\sum_{\omega \in \Omega} X(\omega) \Pr(\omega) &= \sum_x \sum_{\{\omega \in \Omega : X(\omega) = x\}} X(\omega) \Pr(\omega) \\
&= \sum_x \sum_{\{\omega \in \Omega : X(\omega) = x\}} x \Pr(\omega) \\
&= \sum_x x f_X(x).
\end{aligned}
$$

- **Theorem.** $E(X + Y) = E(X) + E(Y)$
- **Proof.**

$$
\begin{aligned}
E(X + Y) &= \sum_{\omega \in \Omega} (X + Y)(\omega) \Pr(\omega) \\
&= \sum_{\omega \in \Omega} [X(\omega) + Y(\omega)] \Pr(\omega) \\
&= \sum_{\omega \in \Omega} X(\omega) \Pr(\omega) + \sum_{\omega \in \Omega} Y(\omega) \Pr(\omega)
\end{aligned}
$$

Examples

- Back to the expected value of tossing two dice:
  - Let X be the count on the first die, Y the count on the second die.

    $$E(X) = E(Y) = (1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$$

  - So

    $$E(X + Y) = E(X) + E(Y) = 3.5 + 3.5 = 7$$

- Back to the expected value of $B_{n,p}$.
  - Consider a sequence of n Bernoulli trials.
  - Let X_k be the indicator function of success in the kth trial.
  - We already know $E(X_k) = p$.
  - $X = \sum_{k=1}^{n} X_k$ is distributed $B_{n,p}$.
  - Therefore

    $$E(X) = E\Big(\sum_{k=1}^{n} X_k\Big) = \sum_{k=1}^{n} E(X_k) = np.$$

    The middle equality can be derived by induction.