Cumulative Distribution Functions

- Let y be a discrete random variable which takes on values $y_1 < y_2 < \cdots$, with respective probabilities $p_1, p_2, \ldots$.

- What is the $P(y \leq y_4)$?

- Since the events $E_i = \{y = y_i\}$ are mutually exclusive,

$$P(y \leq y_4) = \sum_{i=1}^{4} P(y = y_i) = \sum_{i=1}^{4} p_i.$$

- The **cumulative distribution** of y is

$$F(a) = P(y \leq a) = \sum_{i:y_i \leq a} p_i$$

Expected Value

The **mean** or **expected value** of y is the *weighted average of possible values of y*,

$$\mu = \mathbb{E}(y) = \sum_{\text{all y}} yp(y).$$

Some Useful Expectation Theorems

- $\mathbb{E}(c) = c$

- $\mathbb{E}(cy) = c\mathbb{E}(y)$

- $\mathbb{E}(x + y) = \mathbb{E}(x) + \mathbb{E}(y)$ where x and y are two random variables.

- Expectation is linear, i.e. $\mathbb{E}(3x + 5y) = 3\mathbb{E}(x) + 5\mathbb{E}(y)$.

- Let $g(y)$ be a function of y, then the **mean** or **expected value** of $g(y)$ is

$$\mathbb{E}[g(y)] = \sum_{\text{all y}} g(y)p(y).$$

- $\mathbb{E}[g_1(y) + \cdots + g_k(y)] = \mathbb{E}[g_1(y)] + \cdots + \mathbb{E}[g_k(y)]$.

- The **variance** of y is

$$\sigma^2 = \mathbb{E}(y - \mu)^2,$$

which is also equal to

$$\begin{aligned} \sigma^2 &= \mathbb{E}(y^2 - 2y\mu + \mu^2) \\ &= \mathbb{E}(y^2) - 2\mu\mathbb{E}(y) + \mu^2 \\ &= \mathbb{E}(y^2) - 2\mu^2 + \mu^2 \\ &= \mathbb{E}(y^2) - \mu^2 \end{aligned}$$

- The **standard deviation** of y is $s = \sqrt{\sigma^2}$.

Bernoulli Distribution

- **Some examples**

(1) Flip a coin. $y = 1$, if it is a head; 0, otherwise.

(2) You guess at a multiple choice question which has four choices. $y = 1$ if your answer is correct; 0, if it is wrong.

- **Bernoulli Trial**

Sample space = { **S**uccess, **F**ailure }.

$P(S) = p,\ P(F) = q = 1 - p.$

- **Bernoulli random variable** $y = 1$, if a success occurs and $y = 0$ if a failure occurs.

$$p(y) = p^y q^{1-y},$$

where $q = 1 - p$, and $y = 0, 1$.

Mean and Variance of a Bernoulli Random Variable

$$\begin{aligned}\mu &= \sum yp(y) \\ &= 1 \cdot p + 0 \cdot (1-p) \\ &= p\end{aligned}$$

$$\begin{aligned}\sigma^2 &= \mathbb{E}(y^2) - \mu^2 \\ &= [\sum y^2 p(y)] - \mu^2 \\ &= 1 \cdot p + 0 \cdot (1-p) - p^2 \\ &= p - p^2 \\ &= p(1-p) = pq.\end{aligned}$$

Binomial Distribution

- The experiment consists of n **identical** (i.e., $P(S) = p$ and $P(F) = q$ remain the same from trial to trial) and **independent** Bernoulli trials

- The binomial random variable y is the number of successes in n trials.

$$p(y) = \binom{n}{y} p^y q^{n-y} \quad (y = 0, 1, \ldots, n)$$

- $\mu = np$ and $\sigma^2 = npq$.

How to Derive $p(y)$?

Show that $\mu = np$.

If y has a binomial (n, p) distribution, then $y = x_1 + x_2 + \dots x_n$, where each x_i is independent Bernoulli (p).

$$\begin{aligned} \mathbb{E}(y) &= \mathbb{E}(x_1 + x_2 + \cdots + x_n) \\ &= \mathbb{E}(x_1) + \mathbb{E}(x_2) + \cdots + \mathbb{E}(x_n) \\ &= p + \cdots + p \\ &= np \end{aligned}$$

Binomial Examples

- Flip a coin 10 times. $y =$ number of heads observed.
- A multiple choice test contains 10 questions, each with four choices, and you guess at each question. $y =$ the number of questions answered correctly.
- $y =$ number of accidents that occur along I-40 during a one-month period.
- Because not all airline passengers show up for their reserved seat, an airline sells 125 tickets for a flight that holds only 120 passengers. $y =$ number of passengers who show up.
- A particular long traffic light on your morning commute is green 20% of the time that you approach it. Assume that each morning represents an independent trial. $y =$ number of mornings when the light is green, during a week.
- ...

Example 4.8 Electrical engineers recognize that high neutral current in computer power systems is a potential problem. A recent survey of computer power system load currents at U.S. sites found that 10% of the sites had high neutral to full-load current ratios. If a random sample of five computer power systems is selected from the large number of sites in the country, what is the probability that:

(a) Exactly three will have a high neutral to full-current load ratio?

(b) At least three?

(c) If 4 out of 5 computer power systems are found to have high neutral to full-load current ratio, how would you infer the true value of p?

Multinomial Distribution

- The experiment consists of n **identical** and **independent** trials.
- There are k possible outcomes to each trial and the probabilities of the k outcomes are denoted by $p_1, p_2, \ldots, p_k$ and $\sum_{i=1}^{k} p_i = 1$.
- The random variables are the counts $y_1, \ldots, y_k$ in each of the k classification categories, where $\sum_{i=1}^{k} y_i = n$.
- $p(y_1, \ldots, y_k) = \frac{n!}{y_1! \ldots y_k!} p_1^{y_1} \ldots, p_k^{y_k}$
- $\mu_i = np_i$ and $\sigma_i^2 = np_1(1 - p_i)$.

Example 4.13 : A manufacturer uses electrical fuses in an electronic system. The fuses are purchased in large lots and tested sequentially until the first defective fuse is observed. Let y be the number of fuses that need to be tested. Assume that the lot contains 10% defective fuses. Determine the probability distribution of y.

Let 0 denote that a fuse is not defective and let 1 denote that a fuse is defective.

The sample space of the experiment is infinite and it can be represented as all possible binary sequence with a string of 0's and end with 1, i.e.

$$S = \{1, 01, 001, 0001, 00001, \ldots\}$$

$$\begin{aligned}
P(y = 1) &= 0.1 \\
p(y = 2) &= (1 - 0.1) \times 0.1 \\
p(y = 3) &= (1 - 0.1)^2 \times 0.1 \\
&\ldots \\
p(y) &= 0.1(1 - 0.1)^{y-1}
\end{aligned}$$

Geometric Distribution

- Experiment consists a series of identical and independent Bernoulli trials

$$P(S) = p, \quad P(F) = q = 1 - p.$$

Instead of a fixed number of trials, the trials are conducted until a success is obtained.

- The geometric random variable y is the number of trials until the first success is observed.

- $p(y) = (1-p)^{y-1}p$, $y = 1, 2, \ldots$.

  Note that the probability of $y = i$ is $(1-p)$ times the probability of $y = i - 1$. That is, the probabilities decrease in a *geometric* progression.

- $\mu = 1/p$ and $\sigma^2 = (1-p)/p^2$.

Negative Binomial Distribution

- Experiment consists a series of identical and independent Bernoulli trials

- $P(S) = p$ and $P(F) = q = 1 - p$.

- The negative binomial random variable y is the number of trials until the rth success is observed

- $p(y) = \binom{y-1}{r-1} p^r q^{y-r} \quad (y = r, r+1, \ldots)$

- $\mu = r/p$ and $\sigma^2 = rq/p^2$.

Example : A listing of customer accounts at a large corporation contains 1000 customers. Of these, 700 have purchased at least one of the corporation's products in the last three months. To evaluate a new product design, 50 customers are sampled at random from the list. Let y denote the number of sampled customers who have purchased from the corporation in the last three months.

Find the distribution for y.

Hypergeometric Distribution

- The experiment consists of randomly drawing n elements **without replacement** from a set of N elements, r of which are S's and $(N - r)$ of which are F's.

- The hypergeometric random variable y is the number of S's in the draw of n elements,

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}},$$

where $y = \text{Max}[0, n - (N - r)], \ldots, \text{Min}[r, n]$

- $\mu = nr/N$ and $\sigma^2 = \frac{r(N-r)n(N-n)}{N^2(N-1)}$