FALL 2013
Stat 8003: Statistical Methods I
Lecture 3

Jichun Xie

1 Probability

1.1 Sample Spaces

In real world, there are lots of events occur randomly. In Statistics, such events are called *experiments*. The set of all possible outcomes is the *sample space* corresponding to an experiment. The sample space is denoted by Ω and an element of Ω is denoted by ω.

Example 1. We can flip a coin once and record the side faces up. The sample space of this experiment is

$$\Omega = \{\mathrm{H}, \mathrm{T}\}.$$

Example 2. We can flip a coin three times and record every time the side faces up. The sample space of this experiment is

$$\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$$

Example 3. Mary went to a shopping mall. The time she spent there is

$$\Omega = \{t \mid t \geq 0\}$$

We are often interested in particular subsets of Ω, which in probability languages are called *events*. In Example 2, the event that the second flip of the coin is a head is the subset of Ω denoted by

$$A = \{HHH, THH, HHT, THT\}$$

In Example 3, the event that Mary's shopping time is less than 30 minutes is

$$A = \{t \mid 0 \leq t < 30\}$$

1.2 Probability Measures

A probability measure on is a function $\mathtt{P}$ from subsets of Ω to the real numbers that satisfies the following axioms:

1. $\mathtt{P}(\Omega) = 1$.
2. If $A \subset \Omega$, then $\mathtt{P}(A) \geq 0$.
3. If A_1 and A_2 are disjoint, then

$$\mathtt{P}(A_1 \cup A_2) = \mathtt{P}(A_1) + \mathtt{P}(A_2).$$

   More generally, if $A_1, \ldots, A_n$ are mutually disjoint, then

$$\mathtt{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathtt{P}(A_i).$$

Some properties of probability measure:

A. $\mathtt{P}(A^c) = 1 - \mathtt{P}(A)$.

B. $\mathtt{P}(\emptyset) = 0$.

C. If $A \subset B$, then $\mathtt{P}(A) \leq \mathtt{P}(B)$.

D. $\mathtt{P}(A \cup B) = \mathtt{P}(A) + \mathtt{P}(B) - \mathtt{P}(A \cap B)$.

How to compute probability?

Counting. Suppose the coin in Example 2. is a fair coin. Let $A = \{HHH, THH, HHT, THT\}$, *i.e.* the coin's head is up in the second flip. Then $\mathtt{P}(A) = 1/2$. !

Multiplication. If there are p (independent) experiments, each with outcome set Ω_i, $i = 1, \ldots, p$. Suppose $A = A_1 \otimes A_2 \otimes \ldots \otimes A_p$, where $A_i \subset \Omega_i$, $i = 1, \ldots, p$. Then

$$\mathtt{P}(A) = \prod_{i=1}^{p} \mathtt{P}(A_i).$$

1.3 Conditional Probability

Screening test is a commonly seen test in biomedical studies. Patients or subjects receive screening tests to pre-diagnose whether she/he has the disease of interest. Suppose

$$\begin{aligned} D &= \text{person has disease of interest;} \\ N &= \text{person does not have disease;} \\ T+ &= \text{person gives positive test response;} \\ T- &= \text{person gives negative test response;} \end{aligned}$$

Example. Doctor. White sampled 1000 people known to have diabetes, and 1000 known to not have diabetes:

	Disease Status		
Test Result	Present (D)	Absent (N)	Total
$T+$	950	10	960
$T-$	50	990	1040
Total	1000	1000	2000

$$\mathtt{P}(D \mid T+) = 950/960 = 0.98$$
$$\mathtt{P}(T+ \mid D) = 950/1000 = 0.95$$

Question: What is $\mathtt{P}(N \mid T-)$ and $\mathtt{P}(T \mid N)$? How to interprete?

Definition 1. *Let A and B be two events with $\mathtt{P}(B) \neq 0$. The conditional probability of A given B is defined to be*

$$\mathtt{P}(A \mid B) = \frac{\mathtt{P}(A \cap B)}{\mathtt{P}(B)}.$$

Multiplication Law. Let A and B be the events and assume $\mathtt{P}(B) \neq 0$. Then

$$\mathtt{P}(A \cap B) = \mathtt{P}(A \mid B)\mathtt{P}(B).$$

Law of Total Probability. Let $B_1, \ldots, B_n$ be such that $\bigcup_{i=1}^n B_i = \Omega$ and $B_i \cap B_j = \emptyset$ for $i \neq j$, with $\mathtt{P}(B_i) > 0$ for all i. Then for any event A,

$$\mathtt{P}(A) = \sum_{i=1}^{n} \mathtt{P}(A \mid B_i)\mathtt{P}(B_i).$$

1.4 Independence

Definition 2. *A and B are said to be independent events if* $\mathtt{P}(A \cap B) = \mathtt{P}(A)\mathtt{P}(B)$.

How to interpretate independence? We say two events A and B are independent if knowing that one had occurred gave us no information about whether the other had occurred. Let's suppose $\mathtt{P}(B) > 0$. Then

$$\mathtt{P}(A \mid B) = \frac{\mathtt{P}(A \cap B)}{\mathtt{P}(B)} = \mathtt{P}(A).$$

Similarly, if $\mathtt{P}(A) > 0$, $\mathtt{P}(B \mid A) = \mathtt{P}(B)$.

Let's review Example 2. Let A denote the event of head on the first toss, and B the event of heads on the second toss. Write out A and B mathematically. What is $A \cap B$? What is $\mathtt{P}(A \cap B)$?

2 Random Variables

2.1 Discrete Random Variables

2.1.1 Bernoulli Distribution

In Example 1, define

$$X = \text{the count of heads} = \begin{cases} 1 & \text{if head;} \\ 0 & \text{otherwise.} \end{cases}$$

Then X is called a *random variable*, with two possible values; 0 and 1. If it is a fair coin. Then

$$\mathtt{P}(X = 0) = \mathtt{P}(X = 1) = 1/2.$$

Let X = the indicator of a single trial. Then $X \sim \text{Ber}(\theta)$ with

$$\begin{aligned} &\mathtt{P}(X = 1) = \theta \\ &\mathtt{P}(X = 0) = 1 - \theta \\ &\mathtt{P}(X = k) = \theta^k (1 - \theta)^{1-k}, \ y \in \{0, 1\} \end{aligned}$$

2.1.2 Binormial Distribution

In Example 2, also define

$$X = \text{the count of heads.}$$

Since now the coin is flipped three times, the possible values X can take are 0, 1, 2, and 3. Let's study what's the probability X can take these values later.

Let X = the number of success observed in the n trials. Then y Bin(n, p) with

$$\mathtt{P}(X = k) = \binom{n}{\theta}\theta^k(1-\theta)^{n-k},\ k \in \{0, \ldots, n\}$$

2.1.3 Geometric Distribution

In a large traumatic brain injury experiment, put injured and uninjured rats in a Morris water maze and determine whether each animal reaches the platform in 60 seconds. Repeat the experiment until the animal reaches the platform before 60 seconds and record the number of trials up until the first success. Large number of trials suggest greater cognitive injury.

Let X = the number of trials up to and including the first success. Then $X \sim$ Geom(θ), with

$$P(X = k) = \theta(1-\theta)^{k-1},\ k = 1, 2, \ldots$$

2.1.4 Negative Binomial Distribution

The negative binomial distribution is a generalization of the geometric distribution. Suppose that a sequence of independent trials, each with probability of success p, is performed until there are r successes in all. Let X = denote the total number of trials.

$$\mathtt{P}(X = k) = \binom{p-1}{r-1}p^r(1-p)^{k-r},\ k = r, r+1, \ldots$$

2.1.5 Hyptergeometric Distribution

Suppose that an urn contains n balls, of which r are black and $n - r$ are white. Let X denote the number of black balls drawn when taking m balls without replacement. Then $X \sim$ Hyp(r, n, m) with

$$\mathtt{P}(X = k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}},\ \max(0, m-n+r) \leq k \leq \min(r, m)$$

2.1.6 Poisson Distribution

Consider the following examples:

- the number of calls to a suicide hotline on a given 24 hour period;
- the number of cells in a culture that exhibit a genetic mutation;
- the number of deaths during heatwaves in the US.

Let X = the number of independent events per unit time that occur for some rate λ. $X \sim$ Poi(λ), then

$$\mathtt{P}(X = k) = \frac{\lambda^k}{k!}\exp(-\lambda),\ k \in \{0, \ldots, \infty\}$$

2.1.7 Example

In a specific teaching hospital, HUZ, we have 25 Caucasians and 15 African American patients awaiting surgery. Of these, 30 will receive care largely from residents and 10 largely from attending surgeon. If Caucasians and African Americans are equally likely to receive care from a resident, what is the probability that all 15 of the American American patients receive care from resident?

Let X = the number of African American who received care from a resident at HUZ. $X \sim$ $\mathrm{Hyp(n, r, m)}$

$$\begin{aligned}
n &= 40 \text{ (total 40 patents);} \\
r &= 15 \text{ (15 caucations);} \\
m &= 30 \text{ (30 draws = 30 resident treatment spots);}
\end{aligned}$$

$$\mathtt{P}(X \geq 15) = \mathtt{P}(X = 15) = \frac{\binom{r}{15}\binom{n-r}{m-15}}{\binom{n}{15}} = \frac{\binom{15}{15}\binom{25}{15}}{\binom{40}{30}} = .0039$$

It seems that the probability is unusually low.

Now suppose that in teaching hospital more generally, the probability of being treated by a resident is 0.8. How might this information change the setup of the problem?

(i). Was the probability of being treated at HUZ by a resident unusual (particularly high)?

Let $Y = \#$ of patients receiving care from a resident at HUZ. $Y \sim \text{Bin}(40, 0.8)$.

$$P(Y \geq 30) = \sum_{k=30}^{40} \binom{40}{k} 0.8^k (1-0.8)^{40-k} \approx 0.74$$

(ii) Was the probability of being treated by a resident given African American unusual?

Let $Z = \#$ African Americans receive care from a resident at HUZ $Z \sim \text{Bin}(15, 0.8)$.

$$P(Z \geq 15) = P(X = 15) = \binom{15}{15} 0.8^{15} 0.2^0 \approx 0.035.$$

2.2 Continuous Random Variables

2.2.1 Uniform Distribution

Let $X = \text{Unif}(a, b)$. Then the probability density function (pdf) on $[a, b]$ is

$$f(x) = \begin{cases} 1/(b-a), & a \leq x \leq b \\ 0, & x < a \text{ or } x > b \end{cases}$$

Note that for continuous random variable, the probability it takes a specific value is

$$\mathtt{P}(X = c) = \int_c^c f(x)\,\mathrm{d}x = 0.$$

The cumulative density function (cdf) is

$$F(x) = \mathtt{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Here $\mathtt{P}(X \leq x) = \mathtt{P}(X < x)$ since X is a continuous random variable.

2.2.2 Normal Distribution

Let $X \sim \text{N}(\mu, \sigma^2)$, a normal distribution with mean μ and variance σ^2. Then pdf of $\text{N}(\mu, \sigma^2)$ is

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\frac{-(y-\mu)^2}{2\sigma^2}\};\ y \in \mathbb{R}.$$

Normal pdf is a bell-shaped curve over the entire real line.

Let $Z = (Y - \mu)/\sigma$, then $Z \sim \mathrm{N}(0, 1)$. Z is said to follow the standard normal distribution.

Normal distribution is symmetric around μ.

$$P(\mu - \sigma < X < \mu + \sigma) = P(-1 < Z < 1) \approx 0.683$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3) \approx 0.997$$

Examples: Height of children of age 8, IQ, *etc.*.

2.2.3 Exponential Distribution

Let $X \sim \mathrm{Exp}(\lambda)$. Then the pdf and cdf of X are

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} \ (x > 0)$$
$$F(x) = (1 - e^{-x/\lambda}) I(x > 0).$$

Exponential distributions are skewed distributions on the positive real line. $1/\lambda$ is sometimes thought of as an instantaneous failure rate, or a "hazard". This is a special case of the Gamma distribution.Sometimes, reparametralize $\lambda^* = 1/\lambda$. Then $X \sim \mathrm{Exp}(\lambda^*)$ with

$$f(x) = \lambda^* \exp(-\lambda^* x)$$

In R function `dexp, rexp, qexp`, the argument rate $= \lambda*$.

Exponential distribution is sometimes used to characterize failure time, *e.g.* the function time of a light bulb.

2.2.4 Gamma Distribution

For $\alpha > 0$, the Gamma function is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x) \, \mathrm{d}x.$$

Note that if $\alpha > 0$, then

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1).$$

Also, $\Gamma(1) = 1$. Thus if α is a positive integer, $\Gamma(\alpha) = (\alpha - 1)!$

Let $X \sim \mathrm{Gam}(\alpha, \lambda)$. Then the pdf is

$$f(x) = \frac{1}{\lambda^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\lambda}, \ x > 0.$$

Here α is called shape parameter and λ is called scale paramter. $\mathrm{Gam}(1, \lambda) = \mathrm{Exp}(\lambda)$. Sometimes, we reparameterize $\beta = 1/\lambda$.

2.3 Summary

What do we use distributions for?

1. Statistical infrence
2. Making predicitons about samples

Example. I obtained information using random telephone survey sampling methods of 100 four-person families living in West Philadelphia and found that 27 families lived below the poverty line. Is there evidence that the poverty rate in West Philadelphia is higher than the national average of 19.8%?

Let $X = \#$ families living in poverty, and $n = \#$ family sampled. Suppose that the poverty rate in W. Philly is equivalent to the national average. Then $X \sim \text{Bin}(n = 100, p = 0.198)$.

$$\mathtt{P}(X \geq 27) = 1 - \mathtt{P}(x < 26) = 1 - \sum_{k=0}^{26} \binom{100}{k} 0.198^k (1 - 0.198)^{100-k}.$$

If the true poverty level in W. Philly is 19.8% and we sample 100 families, the chance seeing 27 or more families living below the poverty line is about 5%. What do you think of this answer?