Statistics ST2334 Topic 4: Probability Distributions (part b)

2011/2012 Semester 2

Expectation

The Expected Value, Expectation or Mean of a discrete random variable X is defined by

$$\mathbf{E}(X) = \sum_{all \;\; different \;\; x_i} x_i P(X = x_i).$$

Example Suppose we toss an even die and the upper face is recorded X. We have $P(X = k) = 1/6$, and

$$\mathbf{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + ... + 6 \cdot \frac{1}{6} = 3.5$$

Some Properties of Expectation

1. If a and b are constants then $\mathbf{E}(a+bX) = a + b\mathbf{E}(X)$. Proof:

$$\begin{aligned}
\mathbf{E}(a+bX) &= \sum_x (a+bx)\, P(X=x) \\
&= a \sum_x P(X=x) + b \sum_x x P(X=x) \\
&= a + b\mathbf{E}(X)\,.
\end{aligned}$$

2. For any random variables X, Y then $\mathbf{E}(X+Y) = \mathbf{E}(X) + \mathbf{E}(Y)$.
3. $\mathbf{E}(X)$ is the constant which minimises $\mathbf{E}\Big((X-c)^2\Big)$.
4. for any function f, $\mathbf{E}(f(X)) = \sum_{x_i} f(x_i) P(X = x_i)$

Variance

The variance of a r.v. X is the number

$$\begin{aligned}\mathbf{Var}(X) &= \mathbf{E}\left((X-\mathbf{E}(X))^2\right) \\ &= \sum_{x_i}(x_i-\mu)^2 P(X=x_i) \\ \text{Standard Deviation} &= \sqrt{\mathbf{Var}(X)}\end{aligned}$$

Properties of Variance

1. $\mathbf{Var}(X) \geq 0$.
2. If $\mathbf{Var}(X) = 0$, then $P(X = \mathbf{E}(X)) = 1$
3. If a, b constants, $\mathbf{Var}((a + bX)) = b^2\mathbf{Var}(X)$
   Proof:
   $$\begin{aligned} \mathbf{Var}(a + bX) &= \mathbf{E}(a + bX - a - b\mathbf{E}(X))^2 \\ &= b^2\mathbf{E}(X - \mathbf{E}(X))^2 \\ &= b^2\mathbf{Var}(X) \end{aligned}$$
4. $\mathbf{Var}(X) = \mathbf{E}\left(X^2\right) - \mathbf{E}(X)^2$
   Proof:
   $$\begin{aligned} \mathbf{E}(X - \mathbf{E}(X))^2 &= \mathbf{E}\left(X^2 - 2X\mathbf{E}(X) + (\mathbf{E}(X))^2\right) \\ &= \mathbf{E}\left(X^2\right) - 2\mathbf{E}(X)\,\mathbf{E}(X) + \mathbf{E}(X)^2 \\ &= \mathbf{E}\left(X^2\right) - (\mathbf{E}(X))^2 \end{aligned}$$

Chebyshev's Inequality

If a probability distribution has mean μ and standard deviation σ, the probability of getting a value which deviates from μ by at least $k\sigma$ is at most $\frac{1}{k^2}$.

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- Example: Applying $k = 2$, we conclude that for ANY random variable X, there is at most $\frac{1}{2^2} = 25\%$ chance that it is 2 standard deviations or further away from its mean.
- For most distributions, it is unlikely for a random variable to be more than 2 SD's away from its mean. (We will see concrete examples later in the course.)

Back to Binomial

If $X \sim B(n,p)$ then $\mathbf{E}(X) = np$, $\mathbf{Var}(X) = np(1-p)$.

$$
\begin{aligned}
\mathbf{E}(X) &= \sum_{r=0}^{n} r \cdot \binom{n}{r} p^r (1-p)^{n-r} \\
&= 0 \cdot (1-p)^n + \sum_{r=1}^{n} r \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} \\
&= np \sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!(n-r)!} p^{r-1} (1-p)^{n-r} \\
&= np \sum_{r=0}^{n-1} \frac{(n-1)!}{(r)!(n-1-r)!} p^r (1-p)^{n-1-r} \\
&= np \sum_{r=0}^{n-1} \binom{n-1}{r} p^r (1-p)^{n-1-r} \\
&= np.
\end{aligned}
$$

Example: Left-handedness

A variety of studies suggest that 10% of the human population is left-handed. If we choose 500 people randomly and let the X be the number of left-handed people in the sample. Then the distribution of X is binomial distribution with $n = 500$ and $p = 0.1$.

- How many left-handers do we expect in the sample of 500?
- Will you be surprised if we ended up with 65 left-handers in our sample of 500?

Back to Hypergeometric

If

$$P(X = x) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}},$$

then

$$\mathbf{E}(X) = n \cdot \frac{a}{N}, ^{1} \quad \mathbf{Var}(X) = n \cdot \frac{a}{N} \cdot \left(1 - \frac{a}{N}\right) \cdot \frac{N-n}{N-1}$$

- Recall the drawing with or without replacement analogy between Binomial and Hypergeometric.
  - Expectation remains the same.
  - Variance is smaller for without replacement.
  - $\frac{N-n}{N-1}$ is some times called the finite population correction factor.

[1] we will prove in tutorial

Back to Poisson

If $X \sim Poi(\lambda)$, then $\mathbf{E}(X) = \lambda, \mathbf{Var}(X) = \lambda.$

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

Poisson Process

- A continuous time process. (Recall Bernoulli process is a discrete time process)
- We count the number of occurrences within some interval of time.
- Defining properties of a Poisson Process with rate parameter α
  - the expected number of occurrences in an interval of length T is αT
  - there are no simultaeneous occurrences
  - the number of occurrences in disjoint intervals of time are independent
- The number of occurrences in any interval of a Poisson Process has a Poisson distribution.

Example: Mutant Turtles

- Dr Eastman was experimenting with mutations in turtles. He exposed them to high dosage radiation and noted that there were about 1 mutation every 4 minutes.
  Find the probabilities of spotting
  (a) one mutation in 3 minutes;
  (b) at least two mutations in 5 minutes;

Geometric Distribution

- The geometric random variable is the number of Bernoulli trials needed to get one success.
- We write $X \sim Geo(p)$ when the Bernoulli trials have probability of success p. Then,

$$P(X = k) = (1 - p)^{k-1} p$$

for k = 1, 2, 3, ...

x	1	2	...	k	...
$P(X = x)$	p	$(1 - p)p$	...	$(1 - p)^{k-1}p$	...

Expectation and Variance of Geometric Random Variable

Let $X \sim Geo(p)$ and let $q = 1 - p$. Then

$$\mathbf{E}(X) = \frac{1}{p}, \quad \mathbf{Var}(X) = \frac{q}{p^2}.$$

Proof:

$$\begin{aligned}
\mathbf{E}(X) &= \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1} \\
&= p\sum_{k=1}^{\infty} \frac{d}{dq}(q^k) = p\frac{d}{dq}\Big(\frac{1}{1-q}\Big) \\
&= p(1-q)^{-2} = \frac{1}{p}
\end{aligned}$$

Expectation and Variance of Geometric Random Variable

$$
\begin{aligned}
\mathbf{E}(X^2) &= \sum_{k=1}^{\infty} k^2 p q^{k-1} \\
&= p\left(\sum_{k=1}^{\infty} k(k+1)q^{k-1} - \sum_{k=1}^{\infty} k q^{k-1}\right) \\
&= p\left(\frac{2}{(1-q)^3} - \frac{1}{(1-q)^2}\right) = \frac{2}{p^2} - \frac{1}{p}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{Var}(X) &= \mathbf{E}(X^2) - \mathbf{E}(X)^2 \\
&= \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} \\
&= \frac{1-p}{p^2} = \frac{q}{p^2}
\end{aligned}
$$

Geometric Distribution Examples

Example Suppose a drunkard keeps throwing darts until he hits the bullseye. The probability distribution of the number of times it is thrown is supported on the infinite set $\{1, 2, 3, \dots\}$ and is a geometric distribution with $p = P(\text{hit bullseye})$.

Example Recall the Toto addict that buys a System 7 every draw. He keeps telling himself, one day he will win, and after that he can retire comfortably. Just like the dart example above, the probability distribution of the number of draws he has to buy until he wins is a geometric distribution with $p = P(\text{win Toto})$.

Negative Binomial

- Number of trials in a Bernoulli Process needed to obtain r successes.
- Generalizes the Geometric distribution.
- Let X be negative binomial distributed with parameters p, r. Then the probability mass function is given by

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \text{ for } x = r, r+1, \ldots$$

and

$$\mathbf{E}(X) = \frac{r}{p} \text{ and } \mathbf{Var}(X) = \frac{r(1-p)}{p^2}$$

Example: Rolling Sixes

- We keep rolling a fair die.
  (a) What is the probability of rolling the 1st "6" on the 10th roll?
  (b) What is the probability of rolling the 4th "6" on the 10th roll?

Multinomial Distribution

- As the name suggests, it is a generalization of the binomial distribution.
- The assumptions remain the same except that each trial has $k \geq 2$ possible outcomes.
  - the k outcomes are mutually exclusive for each trial
  - the probability of each possible outcome namely $p_1, p_2, \ldots, p_k$ are fixed (and sum to 1).
  - the n trials are independent
- The probability of observing x_1 outcomes of the first kind, $\ldots, x_k$ outcomes of the kth kind, is

$$f(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}, \text{ when } \sum_{i=1}^{k} x_i = n$$

Example: LT27 is hard to get to

To get to LT27, students go by foot, by the NUS internal shuttle bus, or by taxi/car. The proportions are 0.3, 0.6, 0.1 respectively. 10 students are randomly selected from lecture (attended by gazillion students). What is the probability that 4 arrived on foot, 5 arrived by ISB and 1 arrived by taxi/car?

Discrete Distributions Recap

- Bernoulli: success with probability p.
- Binomial: number of successes in n independent Bernoulli trials.
- Hypergeometric: number of successes when drawing without replacement
- Poisson: a common distribution for number of events in a fixed period of time with known average rate.
- Geometric: number of Bernoulli trials until first success.
- Negative Binomial: number of Bernoulli trials until rth success.
- Multinomial: Generalizes Binomial for trials with more than 2 outcomes.