Bernoulli, Binomial and Geometric

Math 477

October 6, 2014

1 The Bernoulli

Definition 1.1. *X is a Bernoulli RV with parameter $p, 0 \leq p \leq 1$, denoted as Bernoulli(p) or Ber(p) if $P(X = 1) = p$ and $P(X = 0) = 1 - p$..*

Remark: The Bernoulli RV models the one trial experiment with success probability p, where 1 represents a success and 0 a failure.

1.1 Expectation and Variance

It is clear that if X is a Bernoulli(p) then $E(X) = p$ and $Var(X) = p - p^2 = p(1-p)$.

2 The Binomial

Definition 2.1. *X is a Binomial RV with parameters n, p, $n \geq 1$ an integer $0 \leq p \leq 1$, denoted as Bin(n,p) if*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Remark: The Binomial random variable models a n trials experiment, where *all trials are independent* and each trial's success probability is p.

We check that the formula above indeed gives a valid distribution:

a. It is clear that $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \geq 0$.

b. From the Binomial theorem,

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

Thus $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \leq 1$, and it is indeed a probability distribution.

2.1 Expectation

Lemma 2.2. *Let X be a Bin(n,p) RV. Then $E(X) = np$.*

Proof. We have

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

since the sum is over the probability distribution of a Bin(n-1,p) RV.

2.2 Variance

Lemma 2.3. *Let X be a Bin(n,p) RV. Then $Var(X) = np(1-p)$.*

Proof. We need to compute $E(X^2)$. Arguing similarly as the above we have

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= n \sum_{k=1}^{n} k \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} p^k (1-p)^{(n-1)-(k-1)} \\
&= n \sum_{k=0}^{n-1} (k+1) \frac{(n-1)!}{k!((n-1)-k)!} p^{k+1} (1-p)^{((n-1)-k} \\
&= np \sum_{k=0}^{n-1} (k+1) \frac{(n-1)!}{k!((n-1)-k)!} p^{k} (1-p)^{((n-1)-k} \\
&= np[(n-1)p+1],
\end{aligned}$$

since the sum is equal to the expectation of $Y+1$ where Y has Bin(n-1,p) distribution.

Thus

$$Var(X) = E(X^2) - E^2(X) = n(n-1)p^2 + np - (np)^2 = np - np^2 = np(1-p).$$

2.3 Examples

Example 2.4. *A communication system consists of* n *components, each of which independently functions with probability* p*. The total system is said to operate effectively if at least one half of its components function.*

a. What is the probability that a 5-component system function effectively?

b. What is the probability that a 3-component system function effectively?

c. For what p is the 5 component system more likely to function effectively than a 3 component system?

Ans:

a. Let X be the number of functioning components in the 5 system. Then X has Bin(5,p) distribution. Thus

$$P(X \geq 3) = \binom{5}{3} p^3 (1-p)^2 + \binom{5}{4} p^4 (1-p) + p^5.$$

b. Let Y be the number of functioning components in the 3 system. Then Y has Bin(3,p) distribution. Thus

$$P(X \geq 2) = \binom{3}{2} p^2(1-p) + p^3.$$

c. 5 system is more likely to function effectively than a 3 system if

$$10p^3(1-p)^2 + 5p^4(1-p) + p^5 > 3p^2(1-p) + p^3,$$

or

$$3(p-1)^2(2p-1) > 0$$

or

$$p > 1/2.$$

Example 2.5. *Screws produced by a company are defective with porbability 0.01. The company sells screws in package of 10 and offers money-back guarantee if more than 1 screw are defective. What is the probability that a package will be refunded?*

Ans: Let X be the number of defective screws in a package. Then X has distribution Bin(0.01, 10). Thus

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - (.99)^1 0 - 10(.01)(.99)^9. \end{aligned}$$

Example 2.6. *(Coupon selection) Each bag of chips contains a hidden coupon. There are 10 different coupons, and suppose the chance of getting coupon from different bags of chips are independent. Let X be the number of bags of chips one opens before collecting all different coupons.*

a. What is $P(X = 5), P(X = 7), P(X = 8)$?

b. What is $P(X = 10)$?

Ans:

a. It is clear that we need to open at least 10 bags of chips to get 10 different coupons. So $P(X = 5) = P(X = 7) = P(X = 8) = 0$.

b. If we let Y_i be the number of ith coupon we get from opening 10 bags of chips, then Y_i has distribution Bin(1/10, 10). Note that the $Y_i, i = 1, \cdots, 10$ are NOT independent, because $\sum_{i=1}^{10} Y_i = 10$. For example, if $Y_3 = 9, Y_4 = 1$ then all the other $Y_i, i \neq 3, 4$ are 0. So while it is true that

$$P(X = 10) = P(Y_1 = 1, Y_2 = 1, \cdots, Y_{10} = 1),$$

we do not know how to handle the above expression.

So instead we compute $P(X > 10)$. Then

$$P(X > 10) = P(Y_i = 0, \text{ for some } i) = P(Y_1 = 0 \text{ or } Y_2 = 0 \cdots\ Y_{10} = 0),$$

and we can use the inclusion-exclusion formula as before.

Note also that

$$P(Y_i = 0) = \binom{10}{0}(1/10)^0(9/10)^{10} = (9/10)^{10},$$

by the Binomial distribution.

If we let Y_{ij} denote the number coupons NOT of type i and j we get in opening 10 bags of chips, then Y_{ij} has distribution Bin(8/10, 10). Thus

$$P(Y_i = 0, Y_j = 0) = P(Y_{ij} = 10) = (8/10)^{10}.$$

Similar argument gives us the rest of the result as discussed before.

2.4 Binomial distribution for a large number of trials - Stirling formula

Example 2.7. *Suppose in a population the probability of being a male is p and being a female is $(1-p)$. The scientist would like to test the hypothesis whetehr $p = 1/2$ or not. A sample of size $n = 2k$ is pooled from the population for a large k (hence large n). What is the probability that we get exactly k males in this sample?*

Ans: If we let X be number of males in the sample then X has distribution $Bin(2k, p)$. Thus

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

Normally this is all we can say. But for large k there is a very nice approximation of the above probability, via the Stirling's formula.

2.4.1 The Stirling formula

Theorem 2.8. *For n large, $n! \approx n^{n+1/2}e^{-n}\sqrt{2\pi}$. More precisely,*

$$\lim_{n\to\infty} \frac{n!}{n^{n+1/2}e^{-n}\sqrt{2\pi}} = 1.$$

Proof. (You can skip this proof) We only sketch the proof here.

Step 1. $n! = \int_0^\infty x^n e^{-x} dx$. Proof by induction on n.

Step 2. Change of variable: $x = n + \sqrt{n}t$. Then

$$\begin{aligned} \int_0^\infty x^n e^{-x} dx &= \int_{-\sqrt{n}}^\infty (n + \sqrt{n}t)^n e^{-(n+\sqrt{n}t)} \sqrt{n} dt \\ &= n^{n+1/2} e^{-n} \int_{-\sqrt{n}}^\infty (1 + \frac{t}{\sqrt{n}})^n e^{-\sqrt{n}t} dt \end{aligned}$$

Step 3. Show that the function $f^n(t)$, defined as

$$\begin{aligned} f^n(t) &= 0, \ \ t < -\sqrt{n} \\ &= (1 + \frac{t}{\sqrt{n}})^n e^{-\sqrt{n}t}, \ \ t \geq -\sqrt{n} \end{aligned}$$

satisfies $f^n(t) \to e^{-t^2/2}$ as $n \to \infty$.

This can be done by showing

$$\log f^n(t) = n \log \left(1 + \frac{t}{\sqrt{n}}\right) - \sqrt{n}t \to -t^2/2$$

for $|t| \leq \frac{\sqrt{n}}{2}$, using the Taylor expansion:

$$\log(1 + x) = x - \frac{x^2}{2} + O(x^3), |x| \leq 1/2.$$

Note: the following argument is wrong as it involves ∞/∞ which is an indeterminate form: Since

$$(1 + \frac{t}{\sqrt{n}})^n = \left[(1 + \frac{t}{\sqrt{n}})^{\sqrt{n}}\right]^{\sqrt{n}},$$

as $n \to \infty$

$$(1 + \frac{t}{\sqrt{n}})^n \approx e^{\sqrt{n}t}.$$

Therfore $(1 + \frac{t}{\sqrt{n}})^n e^{-\sqrt{n}t} \to 1$ as $n \to \infty$.

Step 4. Show that

$$\int_{-\sqrt{n}}^\infty (1 + \frac{t}{\sqrt{n}})^n e^{-\sqrt{n}t} dt \to \int_{-\infty}^\infty e^{-x^2/2} = \sqrt{2\pi}$$

as $n \to \infty$ by DCT.

Differentiate $f^n(t)$ in n gives

$$\frac{d}{dn} f^n(t) = \log \left(1 + \frac{t}{\sqrt{n}}\right) - \frac{\frac{t}{\sqrt{n}}}{2(1 + \frac{t}{\sqrt{n}})} - \frac{t}{2\sqrt{n}}.$$

Apply Taylor expansion on $\log(1+x)$ gives

$$\begin{aligned}
\frac{d}{dn}\log f^n(t) &= \frac{t}{\sqrt{n}}\frac{1+2\frac{t}{\sqrt{n}}}{2(1+\frac{t}{\sqrt{n}})} - \frac{t^2}{2n} - \frac{t}{2\sqrt{n}} + O([\frac{t}{\sqrt{n}}]^3) \\
&= \frac{t}{2\sqrt{n}}\frac{\frac{t}{\sqrt{n}}}{1+\frac{t}{\sqrt{n}}} - \frac{t^2}{2n} + O([\frac{t}{\sqrt{n}}]^3) \\
&= \frac{\frac{t^2}{2n}}{1+\frac{t}{\sqrt{n}}} - \frac{t^2}{2n} + O([\frac{t}{\sqrt{n}}]^3).
\end{aligned}$$

So that if $t<0$ then $\frac{d}{dn}f^n(t)>0$ and $t>0$ then $\frac{d}{dn}f^n(t)<0$ (for n large). Thus $f^n(t)$ can be dominated by

$$\begin{aligned}
g(t) &= e^{-t^2/2},\ \ t<0 \\
&= f^1(t) = (1+t)e^{-t},\ \ t>0.
\end{aligned}$$

2.4.2 Applying Stirling's formula

Back to our example, since the sample size k is large:

$$\begin{aligned}
\binom{2k}{k} = \frac{(2k)!}{k!k!} &\approx \frac{(2k)^{2k+1/2}e^{-2k}\sqrt{2\pi}}{[k^{k+1/2}e^{-k}\sqrt{2\pi}]^2} = \frac{(2k)^{2k+1/2}e^{-2k}}{k^{2k+1}e^{-2k}\sqrt{2\pi}} \\
&= \frac{2^{2k+1/2}}{\sqrt{k}\sqrt{2\pi}} = \frac{4^k}{\sqrt{k\pi}}
\end{aligned}$$

Hence

$$P(X=k) \approx \frac{[4p(1-p)]^k}{\sqrt{k\pi}}.$$

Remark: If $p=1/2$ then $\sqrt{k}P(X=k)$ reduces to $\frac{1}{\sqrt{\pi}}$. For any other value of p, $\sqrt{k}P(X=k)$ is very close to 0 (converging exponentially fast).

3 The Geometric

Definition 3.1. *X is a Geometric RV with parameters p,* $0 \le p \le 1$*, denoted as Geometric(p) if*

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

Remark: The Geometric RV models the number of trials we must conduct until the first success where the success probability is p.

Example 3.2. *An urn containing 8 white and 10 black balls. Balls are selected randomly with replacement until a black one is obtained. What is the probability that*

a. Exactly n draws are needed?

b. At least k draws are needed?

Ans: Let X denote the number of draws until the first black. Then X has distributio Geometric(10/18) = Geometric(5/9). Thus

a.

$$P(X = n) = (4/9)^{n-1}(5/9) = (4^{n-1}5)/9^n.$$

b.

$$\begin{aligned} P(X \geq k) &= 5/9 \sum_{i=k}^{\infty} (4/9)^{n-1} \\ &= 5/9 \frac{(4/9)^{k-1}}{5/9} = (4/9)^{k-1}. \end{aligned}$$