Bernoulli Random Variable

Let A be an event related to the outcomes of some random experiment.

Indicator function for A is defined by

$$I_A(\zeta) = \begin{cases} 0 & \text{if } \zeta \text{ not in } A \\ 1 & \text{if } \zeta \text{ in } A \end{cases} .$$

I_A is a discrete random variable. In this case, $S_{I_A} =$ range of $I_A = \{0, 1\}$.

Let $P[A] = p$, then pmf is $P_I(0) = 1 - p, P_I(1) = p$.

Identify $I_A = 1$ with a "success".

Binomial Random Variable

A random experiment with two possible outcomes ("success" and "failure") is repeated n times.

Let X be the number of times the "success" event A occurs in these n trials, then the range $S_X = \{0, 1, \cdots, n\}$. Let p be the probability of "success", $0 < p < 1$; p remains the same value for every trial. The trials are independent.

Example Let $X =$ number of heads in n tosses of a coin.

Let I_j be the indicator function for event A in the jth trial,

then $X = I_1 + I_2 + \cdots + I_n$. Binomial random variable is the sum of Bernoulli random variables.

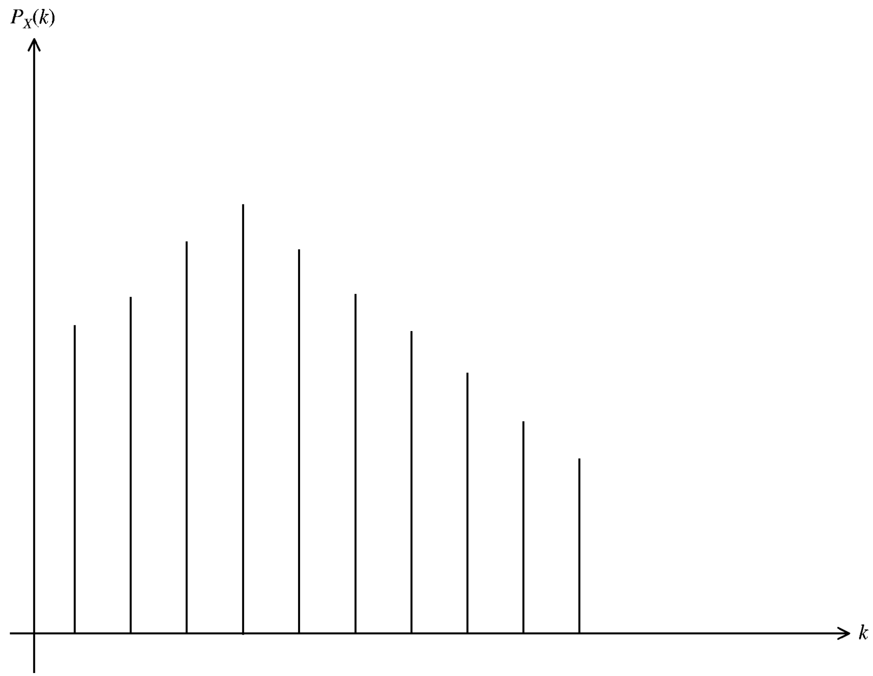
Probability mass function: $P_X(k) = P[X = k] = {}_nC_k p^k (1-p)^{n-k}, k = 0, \cdots, n.$

Most probable number of successes in n trials

$P[X = k]$ is maximum at $k_{max} =$ floor $[(n + 1)p]$, the largest integer less than or equal to $(n + 1)p$. If $(n + 1)p$ is an integer, then the maximum of $P[X = k]$ is achieved at k_{max} and $k_{max} - 1$.

Proof

$P_X(k)$ is seen to be first increasing with k, reaching a maximum value, then decreasing.



Probability mass function of X against the number of successes

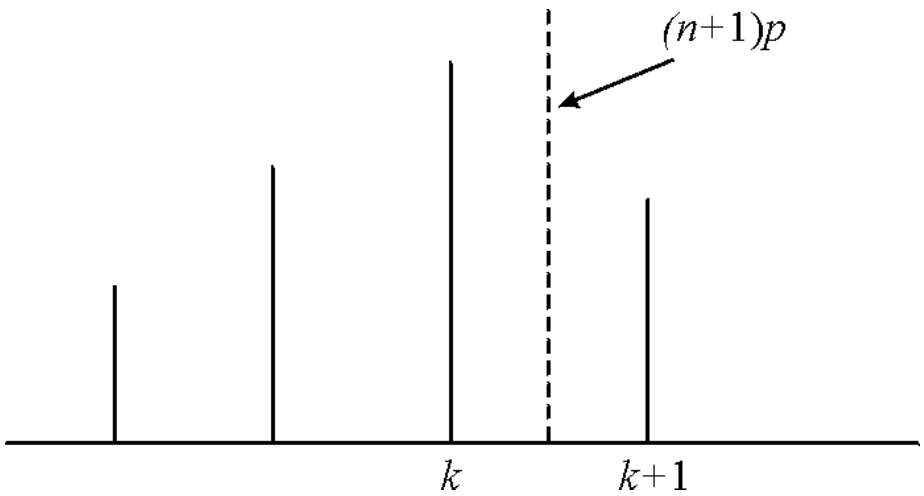
Find the condition on k such that $P_X(k) \geq P_X(k+1)$.

$$ {}_nC_k p^k (1-p)^{n-k} \geq {}_nC_{k+1} p^{k+1} (1-p)^{n-k-1} $$

$$ \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \geq \frac{n!}{(n-k-1)!(k+1)!} p^{k+1} (1-p)^{n-k-1} $$

$$ \frac{1-p}{n-k} \geq \frac{p}{k+1}; \qquad k+1 \geq (n+1)p. $$

The smallest value of k such that $P_X(k) \geq P_X(k+1)$ is floor $[(n+1)p]$, and this is the most probable number of successes.



What happens if $(n+1)p$ is an integer? Write $\widetilde{k}+1=(n+1)p$, then $P_X(\widetilde{k}) = P_X(\widetilde{k}+1)$.

When $k < \widetilde{k}, k < (n+1)p \; -1$ so that $P_X(k) < P_X(k+1)$. When $k > \widetilde{k}+1$, we also obtain $P_X(k) > P_X(k+1)$. The most probable number of successes is either $(n+1)p$ or $(n+1)p-1$.

Example

Suppose the probability of hitting a target is 1/3; and 31 trials of shooting are performed, what is the most likely number of successes?

The good guess is 10, why? Now, $k = \text{floor}\left[\frac{1}{3} \times 32\right] = 10$.

Remark

If the number of trials becomes 32, then the most likely number of successes is either 10 or 11 (both have the same probability value).

Geometric Random Variable

Suppose that independent trials, each having a probability $p, 0 < p < 1$, of being a success, are performed until the first success occurs. If we let X equal the number of trials required, then

$$P[X = n] = (1 - p)^{n-1} p \qquad n = 1, 2, \cdots.$$

It is necessary and sufficient that the first $n - 1$ trials are failures and the nth trial is a success. Also, the outcomes of the successive trials are assumed to be independent.

Since

$$\sum_{n=1}^{\infty} P[X = n] = p \sum_{n=1}^{\infty} (1 - p)^{n-1} = \frac{p}{1 - (1 - p)} = 1,$$

it follows that with probability 1, a success will eventually occur.

Example

An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black one is obtained. If we assume that each selected ball is replaced before the next one is drawn, what is the probability that

(a) exactly n draws are needed;

(b) at least k draws are needed?

Remark

With replacement, the trials are independent and the probability of "success" (a black ball is drawn) remains the same.

Solution

If we let X denote the number of draws needed to select a black ball, then X is a geometric random variable with $p = M/(M+N)$. Hence

(a)

$$P[X=n] = \left(\frac{N}{M+N}\right)^{n-1} \frac{M}{M+N} = \frac{MN^{n-1}}{(M+N)^n}$$

(b)

$$\begin{aligned} P[X \geq k] &= \frac{M}{M+N} \sum_{n=k}^{\infty} \left(\frac{N}{M+N}\right)^{n-1} \\ &= \left(\frac{M}{M+N}\right)\left(\frac{N}{M+N}\right)^{k-1} \bigg/ \left[1 - \frac{N}{M+N}\right] \\ &= \left(\frac{N}{M+N}\right)^{k-1} \end{aligned}$$

The probability that at least k trials are necessary to obtain a success is equal to the probability that the first $k-1$ trials are all failures. That is, for a geometric random variable

$$P[X \geq k] = (1-p)^{k-1}.$$

Memoryless property

The discrete geometric random variable observes the memoryless property:

$$P[X \geq k + j | X > j] = P[X \geq k] \text{ for all } j, k \geq 1.$$

If a success has not occurred in the earlier j trials, then the probability of having to perform at least k more trials to get a success is the same as the probability of initially having to perform at least k trials to get a success.

Proof

First, observe that

$$P[X \geq k] = q^{k-1}, \quad q = 1 - p.$$

We then have

$$P[X \geq k + j | X > j] = \frac{P[X \geq k + j]}{P[X > j]} = \frac{q^{k+j-1}}{q^j} = q^{k-1}.$$

Expected Value and Variance of Discrete Random Variables

Mean or expected value

$$E[X] = \sum_k x_k P_X(x_k).$$

Variance and standard deviation

Extent of the variation of the random variable about its mean

$$\text{VAR}[X] = E[(X - E[X])^2] = E[(X^2 - 2E[X]X + E[X]^2)].$$

Recall that $E[X^2] = \sum_k x_k^2 P_X(x_k)$; $E[E[X]X] = E[X]E[X]$ since $E[X]$ is a fixed quantity, independent of the summing index k.

$$\begin{aligned} \text{VAR}[X] &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2 \\ \text{STD}[X] &= \sqrt{\text{VAR}[X]}. \end{aligned}$$

Expected Value of the Geometric Random Variable

$$E[X] = \sum_{k=1}^{\infty} kpq^{k-1} = p \sum_{k=1}^{\infty} kq^{k-1}.$$

Can we find a closed form for the above summed series?

Recall

$$\begin{aligned} \frac{1}{1-x} &= 1 + x + x^2 + x^3 + \cdots, \text{ for } |x| < 1; \\ \frac{1}{(1-x)^2} &= 1 + 2x + 3x^2 + \cdots, \text{ for } |x| < 1. \end{aligned}$$

Hence, $E[X] = p\left(\frac{1}{1-q}\right)^2 = \frac{1}{p}$.

Example The chance of getting "6" in the throw of a dice is $1/6$. The expected number of trials required to get the first "6" is $\frac{1}{1/6} = 6$. Does the answer sound reasonable?

Variance of the Geometric Random Variable

Using $$\frac{2}{(1-x)^3} = \sum_{k=1}^{\infty} k(k-1)x^{k-2}, \text{ for } |x| < 1.$$

Setting $x = q$ and multiplying both sides by pq, we obtain

$$\frac{2q}{(1-q)^2} \;=\; \sum_{k=1}^{\infty} k^2pq^{k-1} - \sum_{k=1}^{\infty} kpq^{k-1} = E[X^2] - E[X].$$

Since $E[X] = \frac{1}{p}$ so $E[X^2] = \frac{2q}{(1-q)^2} + \frac{1}{p} = \frac{1+q}{p^2}$.

Therefore, $\text{VAR}[X] = E[X^2] - E[X]^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$.

Mean and variance of the binomial random variable

Let X be a binomial random variable with probability of success p and number of trials n. Write $q = 1 - p =$ probability of failure in each trial. The pmf of X is

$$P_X(k) = {}_nC_k p^k q^{n-k}.$$

$$\begin{aligned} E[X] = \sum_{k=0}^{n} k P_X(k) &= \sum_{k=0}^{n} k\, {}_nC_k p^k q^{n-k} = \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} p^k q^{n-k} \\ &= np \sum_{k=1}^{n} \frac{(n-1)!}{[(n-1)-(k-1)]!(k-1)!} p^{k-1} q^{(n-1)-(k-1)}. \end{aligned}$$

Let $n' = n - 1$ and $k' = k - 1$, then

$$E[X] = np \sum_{k'=0}^{n'} \frac{n'!}{(n'-k')!k'!} p^{k'} q^{n'-k'} = np.$$

In a similar manner, by showing that

$$
\begin{aligned}
&\sum_{k=0}^{n} k(k-1)_n C_k q^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(n-k)!(k-2)!} p^{k-2} q^{(n-2)-(k-2)} \\
&= n^2p^2 - np^2,
\end{aligned}
$$

we obtain

$$
\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - E[X]^2 \\
&= \sum_{k=0}^{n} k^2{}_n C_k p^k q^{n-k} - n^2p^2 \\
&= \sum_{k=0}^{n} k(k-1)_n C_k p^k q^{n-k} + \sum_{k=0}^{n} k_n C_k p^k q^{n-k} - n^2p^2 \\
&= (n^2p^2 - np^2) + np - n^2p^2 = npq, \quad q = 1-p.
\end{aligned}
$$