Recall. Responce (N) must assigns distribution →

Some discrete distributions

Definition. Bernoulli distribution - $B(p)$

A Bernoulli distribution takes on only two values: 0 and 1, with probabilities $1-p$ and p, respectively.

- **pmf:** $p(x) = \begin{cases} p^x(1-p)^{(1-x)}, & \text{if } x = 0 \text{ or } x = 1 \\ 0, & \text{otherwise} \end{cases}$
- **mgf:** $pe^t + 1 - p$
- **mean:** p
- **variance:** $p(1-p)$
- **parameter:** $p \in [0, 1]$
- **example:** toss a coin once, p=probability that head occurs

Note: If A is an event, then the indicator random variable I_A follows the Bernoulli distribution.

$\to p = P(A)$

$I_A: \Omega \to \mathbb{R}$, $I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A \end{cases}$

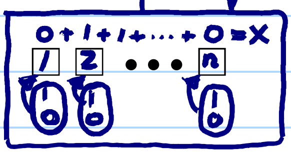
Definition. Binomial distribution - $B(n, p)$

Suppose that n independent Bernoulli trials are performed, where n is a fixed number. The total number of 1 appearing in the n trials follows a binomial distribution with parameters n and p.

shape

explanation

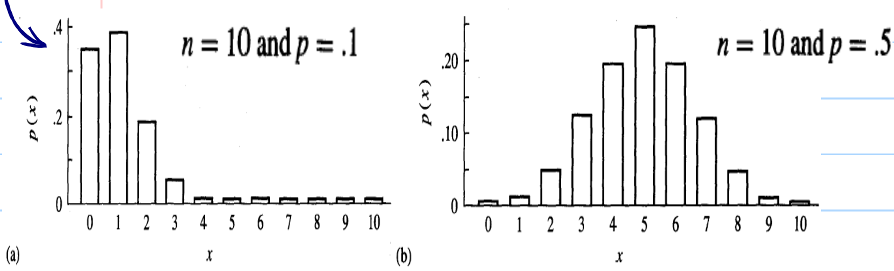
- **pmf:** $p(x) = \begin{cases} \binom{n}{x} p^x(1-p)^{(n-x)}, & x = 0, 1, \ldots, n \\ 0, & \text{otherwise} \end{cases}$



- **mgf:** $(pe^t + 1 - p)^n$, $t \in \mathbb{R}$.
- **mean:** np

intuition

- **variance:** $np(1-p)$ ← max at $p=1/2$, min at $p=0$ or 1
- **parameter:** $p \in [0, 1]$, $n = 1, 2, \ldots$
- **example:** # of heads, toss a coin n times



$B(n,p) \approx N(np, np(1-p))$ as $n \to \infty$.

Note: $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$.

Note.

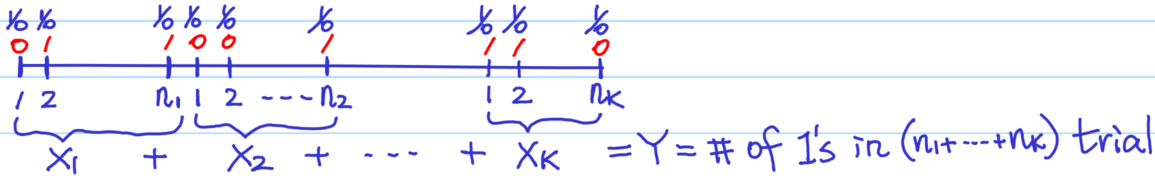
1. binomial distribution is a generalization of bernoulli distribution from 1 trial to n trials

intuition

2. Let $X_1, \ldots, X_n$ be i.i.d. $B(p)$, then $Y = X_1 + \cdots + X_n \sim B(n, p)$.

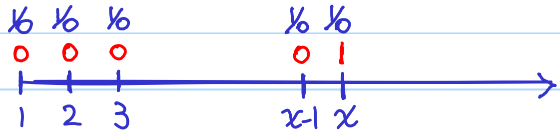
intuition

3. Let $X_i \sim B(n_i, p), i = 1, \ldots, k$, and $X_1, \ldots, X_k$ are independent. Then, $Y = X_1 + \cdots + X_k \sim B(n_1 + \cdots + n_k, p)$.



Definition. Geometric distribution - $G(p)$

The geometric distribution is constructed from an infinite sequence of independent Bernoulli trials. Let X be the total number of trials up to and including the first appearance of 1. Then, X follows the geometric distribution.



explanation

- **pmf:** $p(x) = \begin{cases} (1-p)^{(x-1)}p, & x = 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases}$
- **cdf:** $F(x) = \begin{cases} 1-(1-p)^{[x]}, & 1 \leq [x] \leq x < [x]+1 \\ 0, & x < 1 \end{cases}$
- **mgf:** $\frac{pe^t}{1-(1-p)e^t}$, $t < -\log(1-p)$.
- **mean:** $\frac{1}{p}$

intuition

- **variance:** $\frac{1-p}{p^2}$
- **parameter:** $p \in [0, 1]$
- **example:** lottery, # of tickets a person must purchase up to and including the first winning ticket

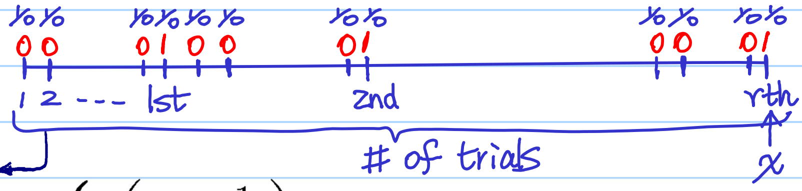
Note: a memoryless distribution ← intuition

$P(X > m+n \mid X > m) = P(X > n)$

Note: $\sum_{x=n}^{\infty} t^x = \frac{t^n}{1-t}$, for $-1 < t < 1$.

Definition. Negative Binomial distribution - $NB(r, p)$

An infinite sequence of independent Bernoulli trials is performed until the appearance of the rth 1. Let X denote the total number of trials. Then, X follows negative binomial distribution.



explanation

- **pmf:** $p(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{(x-r)}, & x = r, r+1, \ldots \\ 0, & \text{otherwise} \end{cases}$
- **mgf:** $\frac{p^r e^{rt}}{[1-(1-p)e^t]^r}$, $t < -\log(1-p)$.
- **mean:** $\frac{r}{p}$

intuition

- **variance:** $\frac{r(1-p)}{p^2}$
- **parameter:** $p \in [0, 1]$, $r = 1, 2, \ldots$
- **example:** lottery, # of tickets a person must purchase up to and including the rth winning ticket

alternative negative binomial

$Y = X - r$

# of 0's when stop — # of trials when stop

Note: $\sum_{x=0}^{\infty} \binom{n+x-1}{x} t^x = \frac{1}{(1-t)^n}$, for $-1 < t < 1$.

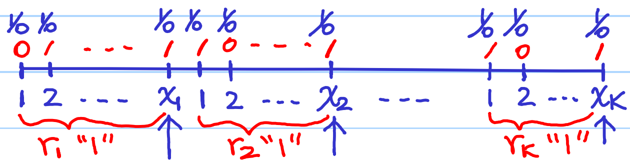
Note.

1. negative binomial distribution is a generalization of geometric distribution from 1st success to rth success

intuition

2. Let $X_1, X_2, \ldots, X_r$ be i.i.d. $G(p)$, then $Y = X_1 + \cdots + X_r \sim NB(r, p)$.
3. Let $X_i \sim NB(r_i, p), i = 1, \ldots, k$, and $X_1, \ldots, X_k$ are independent. Then, $Y = X_1 + \cdots + X_k \sim NB(r_1 + \cdots + r_k, p)$.

intuition



$X_1 + X_2 + \cdots + X_K = Y$ = # of trial until $(r_1 + \cdots + r_K)$ one.

Poisson-Gamma mixture model

4. Consider two random variables (Y, Λ). Suppose that $Y|_{\Lambda=\lambda} \sim Poisson(\lambda)$ and $\Lambda \sim gamma(\alpha, \beta)$, where α is a positive integer. Then,

joint

conditional

marginal

pdf $= \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$

marginal

$$Y + \alpha \sim NB\left(\alpha, \frac{\beta}{1+\beta}\right).$$

# of 0's before observing the αth 1.

Definition. Multinomial distribution - *Multinomial*$(n, p_1, p_2, \ldots, p_r)$

Suppose that each of n independent trials can result in one of r types of outcomes, and that on each trial the probabilities of the r outcomes are $p_1, p_2, \ldots, p_r$. Let X_i be the total number of outcomes of type i in the n trials, $i = 1, \ldots, r$. Then, $(X_1, \ldots, X_r)$ follows a multinomial distribution.

- **joint pmf:**

$$p(x_1, \ldots, x_r) = \begin{cases} \binom{n}{x_1 \cdots x_r} p_1^{x_1} \cdots p_r^{x_r}, & x_i = 0, 1, \ldots, n, \text{ and } \sum_{i=1}^r x_i = n \\ 0, & \text{otherwise} \end{cases}$$

explanation

- **joint mgf:** $(p_1 e^{t_1} + \cdots + p_r e^{t_r})^n$, $t_1, \ldots, t_r \in \mathbb{R}$.
- **marginal distribution:** $X_i \sim B(n, p_i)$, $i = 1, \ldots, r$ — intuition
- **mean:** $E(X_i) = np_i$, $i = 1, \ldots, n$
- **variance:** $Var(X_i) = np_i(1 - p_i)$, $i = 1, \ldots, n$
- **covariance:** $Cov(X_i, X_j) = -np_ip_j$, $i \neq j$ — why negative?
- **parameter:** $p_i \in [0, 1]$, and $\sum_{i=1}^r p_i = 1$. $n = 1, 2, \ldots$

- **example:** randomly choose n people, record the numbers of people with different religions

Note: $(a_1 + \cdots + a_k)^n = \sum_{x_1 + \cdots + x_k = n} \binom{n}{x_1, \cdots, x_k} a_1^{x_1} \cdots a_k^{x_k}$.

Notes.

1. Multinomial distribution is a generalization of the binomial distribution from 2 outcomes to r outcomes.

2. Consider $(X_1, \ldots, X_r) \sim$ *multinomial*$(n, p_1, \ldots, p_r)$. Let $i_0, i_1, \ldots, i_k$ be integers such that $0 = i_0 < i_1 < \cdots < i_k = r$, and define

$$Y_j = \sum_{m=i_{j-1}-1}^{i_j} X_m,$$

$j = 1, \ldots, k$. Then,

$$(Y_1, \ldots, Y_k) \sim \textit{multinomial}(n, q_1, \ldots, q_k),$$

where $q_j = \sum_{m=i_{j-1}-1}^{i_j} p_m$, $j = 1, \ldots, k$.

$(Y_1, Y_2, \ldots, Y_k)$

$\oplus \quad \oplus \quad \cdots \quad \oplus$

X_m's $\quad X_m$'s $\quad \cdots \quad X_m$'s

$\{1, 2, \ldots, \cdot, \ldots, \cdot, r\}$

$i_1 \quad i_2 \quad i_{k-1}$

Definition. Poisson distribution - $P(\lambda)$

Limit of binomial distributions $X_n \sim B(n, p_n)$, where $p_n \to 0$ as $n \to \infty$ in such a way that $\lambda_n \equiv np_n \to \lambda$.

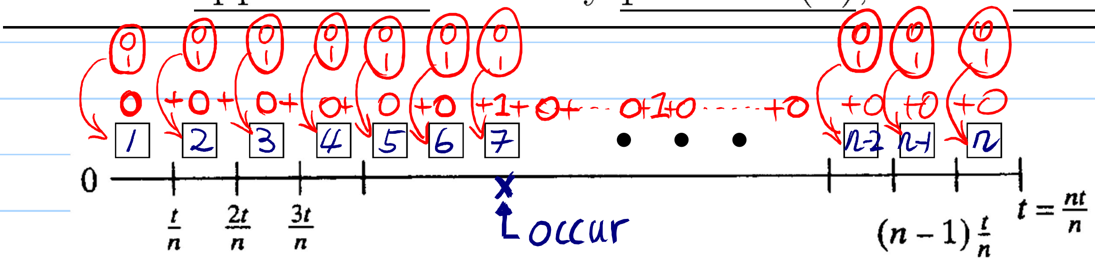
$P_n = \frac{\lambda_n}{n}$

Note: if $a_n \to a$, $\left(1+\frac{a_n}{n}\right)^n \to e^a$.

$$\binom{n}{x} p_n^x (1-p_n)^{(n-x)}$$
$$= \frac{n(n-1)\cdots(n-x+1)}{x!}\left(\frac{\lambda_n}{n}\right)^x\left(1-\frac{\lambda_n}{n}\right)^{n-x}$$
$$= \frac{n(n-1)\cdots(n-x+1)}{n^x}\frac{1}{x!}\lambda_n^x\left(1-\frac{\lambda_n}{n}\right)^{n-x}$$
$$= \underbrace{1}_{1}\underbrace{(1-\frac{1}{n})}_{1}\cdots\underbrace{(1-\frac{x-1}{n})}_{1}\underbrace{\frac{\lambda_n^x}{x!}}_{\frac{\lambda^x}{x!}}\underbrace{\left(1-\frac{\lambda_n}{n}\right)^n}_{e^{-\lambda}}\underbrace{\left(1-\frac{\lambda_n}{n}\right)^{-x}}_{1} \longrightarrow 1^x \cdot \frac{\lambda^x}{x!}\cdot e^{-\lambda}\cdot 1 = \frac{\lambda^x e^{-\lambda}}{x!}$$

explanations.

1. if n large, the pmf of $B(n,p)$ is not easily calculated. Then, we can approximate them by pmf of $P(\lambda)$, where $\lambda = np$.



2. Let X be the number of times some event occurs in a given time interval I. Divide the interval into many small subintervals I_k, $k = 1, \ldots, n$, of equal length. Let N_k be the number of events occurring in I_k. When we can assume $N_1, \ldots, N_n$ are independent and approximately $\sim B(p)$, X has a distribution near $P(\lambda)$, where $\lambda = np$.

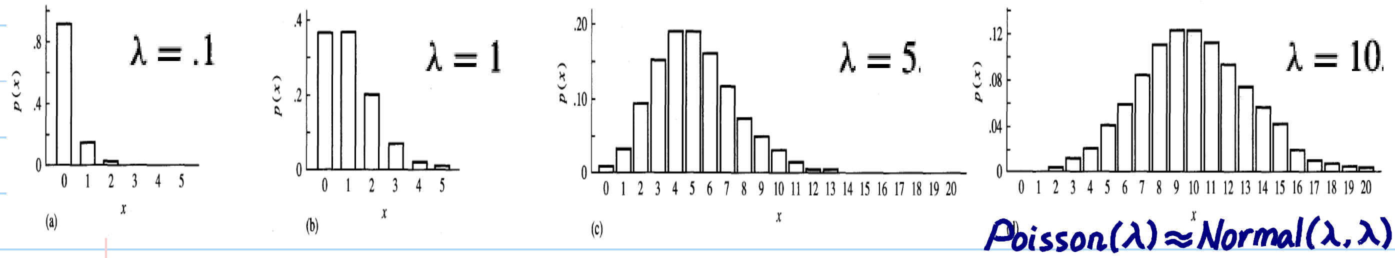
$X = N_1+N_2+\cdots+N_n \sim B(n,p)$ with large n & small p

shape

- **pmf:** $p(x) = \begin{cases} \frac{\lambda^x}{x!}e^{-\lambda}, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$
- **mgf:** $e^{\lambda(e^t-1)}$, $t \in \mathbb{R}$.
- **mean:** λ
- **variance:** λ
- **parameter:** $\lambda > 0$
- **example:** number of phone calls coming into an exchange during a unit of time

meaning of parameter λ: average occurences

Note: $e^\lambda = \sum_{x=0}^{\infty}\frac{\lambda^x}{x!}$.



Poisson(λ) $\approx$ Normal(λ, λ) when λ is large

Notes.

Let $X_i \sim P(\lambda_i)$, $i = 1, \ldots, k$, and $X_1, \ldots, X_k$ are independent. Then,

1. $Y = X_1 + \cdots + X_k \sim P(\lambda \equiv \lambda_1 + \cdots + \lambda_k)$.

   $= Y$ (if know $Y = n$)

   $X_1 + X_2 + X_3 + \cdots + X_k$ ← intuition

   $t_1 \quad t_2 \quad t_3 \quad t_4$ ⇐ $\lambda_1 = \lambda(t_2 - t_1)$, $\lambda_2 = \lambda(t_3 - t_2)$, $\cdots$, → $\boxed{\lambda_1 + \cdots + \lambda_k = \lambda(t_k - t_1)}$

2. $(X_1, \ldots, X_k \,|\, Y = n) \sim multinomial(n, p_1, \ldots, p_k)$, where

$$p_i = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_k} = \frac{\lambda_i}{\lambda}, \; i = 1, \ldots, k.$$

The converse statement also holds with $\lambda_i = \lambda \times p_i$.

Definition. Hypergeometric distribution - $HG(r, n, m)$

Suppose that an urn contains n black balls and m white balls. Let X denote the number of black balls drawn when taking r balls without replacement. Then, X follows hypergeometric distribution. c.f. ⇔ with replacement ⇒ $X \sim B(r, \frac{n}{m+n})$

explanation

- **pmf:** $p(x) = \begin{cases} \dfrac{\binom{n}{x}\binom{m}{r-x}}{\binom{n+m}{r}}, & x = 0, 1, \ldots, \min(r, n),\; r - x \le m \\ 0, & \text{otherwise} \end{cases}$

Note: $\binom{n+m}{r} = \sum_x \binom{n}{x}\binom{m}{r-x}$.

- **mgf:** exist, but no simple expression
- **mean:** $\frac{rn}{n+m}$ (intuition)
- **variance:** $\frac{rnm(n+m-r)}{(n+m)^2(n+m-1)}$
- **parameter:** $r, n, m = 1, 2, \ldots$, and $r \le n + m$
- **example:** sampling industrial products for defect inspection

Notes. a relationship between hypergeometric and binomial distributions: Let $m, n \to \infty$ in such a way that

$$p_{m,n} \equiv \frac{n}{m+n} \to p,$$

where $0 < p < 1$. Then,

intuition: When m, n are large, with replacement ≈ without replacement

$$\frac{\binom{n}{x}\binom{m}{r-x}}{\binom{n+m}{r}} \to \binom{r}{x} p^x (1-p)^{r-x}.$$

optional

required

Reading: Agresti (2013), 1.2

Further Reading: Agresti (2013), 1.3, 1.4, 1.5, 1.6. These are about uni-variate analysis for data from some discrete distribution.

LNp.2-4