Types of variables

- response-explanatory distinction (dependent-independent, response-predictor)
  - ➢ response variables Y: regarded as *random*.
  - ➢ explanatory variables $X_1, \ldots, X_m$: regarded as *deterministic*.
  - ➢ causal relationship? not necessary.
- continuous-discrete distinction
  - ➢ whether a variable can take any values within an interval (if yes ⇒ continuous; if no, only countable values ⇒ discrete)
  - ➢ Q: does continuous *data* really exist in the real world? No, because of the precision limitation of measuring instrument.
  - ➢ a better distinction approach from the viewpoint of data analysis: according the *number* of values a variable can take within the range in which most data fall

    - ▪ variable taking lots of values (dense) ⇒ continuous; few values ⇒ discrete
    - ▪ Q: Should Poisson data (infinite possible values) be treated as discrete or continuous from this data analysis viewpoint?
- quantitative-qualitative distinction
  - ➢ continuous variable must be quantitative
  - ➢ discrete variable could be quantitative/qualitative/between
- categorical (qualitative or between) variables can be further classified into:
  - ➢ nominal variable: no natural ordering between categories (e.g., religious affiliation, mode of transportation, favorite type of music, …)
    - ▪ values that represent categories have no numeric meaning
    - ▪ no value exist between categories
    - ▪ analysis irrelevant to the order of listing the categories

- ordinal variable: there exist some ordering between categories (e.g., size of automobile, social class, political philosophy, patient condition, …)
  - exact distance between ordered categories are unknown
- interval variable: categories are non-overlapping intervals (e.g., functional life length of television set, length of prison term, …)
  - can have numerical distances between 2 categories
  - sometimes, possible to compare the ratio of 2 categories
- Sometimes, it is the way that a variable is measured determined its classification, e.g., education:
  - nominal when measured as public/private school
  - ordinal when measured as none/high school/bachelor/…
  - interval variable when measured by # of year intervals

- hierarchy of measurement scale: interval variable (highest) > ordinal > nominal (lowest)
  - statistical methods for variables of one type can be used for variables at higher level, but not at lower levels
- nominal variable ⇒ qualitative; interval variable ⇒ (close to) quantitative; ordinal variable ⇒ between (fuzzy)

	Resp.	Expl.
Cont.		
Disc. Quan.		
Disc. Inte.		
Ordi.		
Nomi.		

• choice of statistical method/model for different types of variables – a rough classification:

- only response variables, no explanatory variable
  - 1 response variable ⇒ uni-variate analysis
  - more than 1 response ⇒ multi-variate analysis
- both response and explanatory variables

- response: (1) regarded as random variable; (2) modeling depends on the types of variable
  - continuous & normal ⇒ linear model (LM)
  - continuous but not normal (including exponential family, such as Weibull, gamma, …) ⇒ generalized linear model
  - discrete ⇒ generalized linear model (GLM)
- explanatory: (1) regarded as deterministic; (2) same treatment (sum of base functions multiplied by their coefficients) for any types of explanatory variables in modeling
  - quantitative: base functions like polynomial or other continuous transformations (e.g., log, exp, sin, cos, …)
  - qualitative: base function like dummy variables

Sufficient statistics of categorical responses

- **Q**: how to convert categorical observations (symbols or notations) into numerical and computable data without losing any important information? **Q**: what important information?

- Example 1.
  - scenario
    - observe $Y_{\mathbf{x},i} \in r$ categories $\{1, 2, \ldots, r\}$, i=1, …, $k_{\mathbf{x}}$.
    - category j observed with probability $p_{\mathbf{x},j}$, j=1, …, r, and

$$p_{\mathbf{x},1} + p_{\mathbf{x},2} + \cdots + p_{\mathbf{x},r} = 1.$$

    - statistical modeling: $Y_{\mathbf{x},1}, Y_{\mathbf{x},2}, \ldots, Y_{\mathbf{x},k_\mathbf{x}}$ are independent and identically distributed from *multinomial*$(1; p_{\mathbf{x},1}, \ldots, p_{\mathbf{x},r})$.
  - sufficient statistics
    - the joint pmf of $Y_{\mathbf{x},1}, Y_{\mathbf{x},2}, \ldots, Y_{\mathbf{x},k_\mathbf{x}}$ is

$$p_{x,1}^{N_{x,1}} \cdot p_{x,2}^{N_{x,2}} \cdots p_{x,r}^{N_{x,r}}$$

    where $N_{\mathbf{x},j}$ = number of category j in the $k_\mathbf{x}$ trials, j=1, …, r, and $k_\mathbf{x}=N_{\mathbf{x},1}+\cdots+N_{\mathbf{x},r}$.
    - by factorization thm, $(N_{\mathbf{x},1}, \ldots, N_{\mathbf{x},r})$ are the sufficient statistics for the parameters $(p_{\mathbf{x},1}, \ldots, p_{\mathbf{x},r})$ and

$$(N_{\mathbf{x},1}, \ldots, N_{\mathbf{x},r}) \sim multinomial(k_\mathbf{x}; p_{\mathbf{x},1}, \ldots, p_{\mathbf{x},r}).$$

  - note: two categories and binomial is a special case of r=2.

- Example 2.
  - scenario
    - observe $Y_{\mathbf{x},i} \in r$ categories $\{1, 2, \ldots, r\}$, i=1, …, $K_{\mathbf{x}}$, where $K_{\mathbf{x}}$ is a random variable on the non-negative integers.
    - assume $K_{\mathbf{x}} \sim Poisson(\lambda_{\mathbf{x}})$ and given (conditioned on) $K_{\mathbf{x}}=k_{\mathbf{x}}$,
      - $Y_{\mathbf{x},1}, Y_{\mathbf{x},2}, \ldots, Y_{\mathbf{x},k_{\mathbf{x}}}$ are independent and identically distributed from $multinomial(1; p_{\mathbf{x},1}, \ldots, p_{\mathbf{x},r})$, where
      - $p_{\mathbf{x},j}=P$(observing category j), j=1,…, r, and $p_{\mathbf{x},1}+\cdots+p_{\mathbf{x},r}=1$.
  - sufficient statistics
    - the joint pmf of $Y_{\mathbf{x},1}, Y_{\mathbf{x},2}, \ldots, Y_{\mathbf{x},K_{\mathbf{x}}}$ is

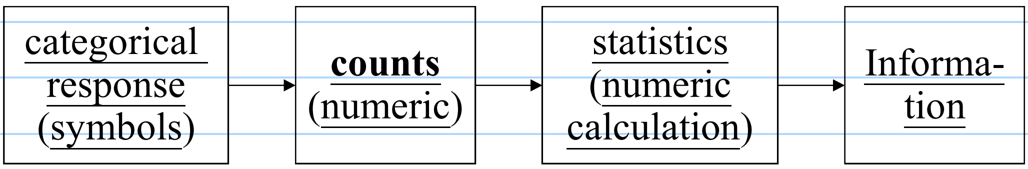
$$\frac{e^{-\lambda_{\boldsymbol{x}}}\lambda_{\boldsymbol{x}}^{K_{\boldsymbol{x}}}}{K_{\boldsymbol{x}}!} \times p_{\boldsymbol{x},1}^{N_{\boldsymbol{x},1}} \cdots p_{\boldsymbol{x},r}^{N_{\boldsymbol{x},r}} \propto \frac{e^{-\lambda_{\boldsymbol{x}} p_{\boldsymbol{x},1}}(\lambda_{\boldsymbol{x}} p_{\boldsymbol{x},1})^{N_{\boldsymbol{x},1}}}{N_{\boldsymbol{x},1}!} \times \cdots \times \frac{e^{-\lambda_{\boldsymbol{x}} p_{\boldsymbol{x},r}}(\lambda_{\boldsymbol{x}} p_{\boldsymbol{x},r})^{N_{\boldsymbol{x},r}}}{N_{\boldsymbol{x},r}!}$$

where $N_{\mathbf{x},j}$ = number of category j in the $K_{\mathbf{x}}$ trials, j=1, …, r, and $K_{\mathbf{x}}=N_{\mathbf{x},1}+\cdots+N_{\mathbf{x},r}$.

    - by factorization thm, $(N_{\mathbf{x},1}, \ldots, N_{\mathbf{x},r})$ are the sufficient statistics for the parameters $(\lambda_{\mathbf{x}} p_{\mathbf{x},1}, \ldots, \lambda_{\mathbf{x}} p_{\mathbf{x},r})$ and

    - $N_{\mathbf{x},1}, \ldots, N_{\mathbf{x},r}$ are independent and $N_{\mathbf{x},j} \sim Poisson(\lambda_{\mathbf{x}} p_{\mathbf{x},j})$.

- In the 2 examples,



Some real-data examples (from Agresti, *Categorical Data Anslysis*)

- From Agresti (2002, 2rd ed.)

TABLE 3.1 Swedish Study on Aspirin Use and Myocardial Infarction

	Myocardial Infarction		
	Yes	No	Total
Placebo	28	656	684
Aspirin	18	658	676

- From Agresti (2013, 3rd ed.)

Table 6.9 Clinical Trial Relating Treatment to Response for Eight Centers, with Expected Value and Variance (of Success Count for Drug) Under Conditional Independence

		Response				
Center	Treatment	Success	Failure	Odds Ratio	μ_{11k}	$\text{var}(n_{11k})$
1	Drug	11	25	1.19	10.36	3.79
	Control	10	27			
2	Drug	16	4	1.82	14.62	2.47
	Control	22	10			
3	Drug	14	5	4.80	10.50	2.41
	Control	7	12			
4	Drug	2	14	2.29	1.45	0.70
	Control	1	16			
5	Drug	6	11	∞	3.52	1.20
	Control	0	12			
6	Drug	1	10	∞	0.52	0.25
	Control	0	10			
7	Drug	1	4	2.0	0.71	0.42
	Control	1	8			
8	Drug	4	2	0.33	4.62	0.62
	Control	6	1			

- From Agresti (2013, 3rd ed.)

Table 2.11 Data for Exercise 2.15 on Graduate Admissions

	Whether Admitted			
	Male		Female	
Department	Yes	No	Yes	No
A	512	313	89	19
B	353	207	17	8
C	120	205	202	391
D	138	279	131	244
E	53	138	94	299
F	22	351	24	317
Total	1198	1493	557	1278

- From Agresti (2013, 3rd ed.)

Table 2.5 Cross-Classification of Smoking by Lung Cancer

	Lung Cancer	
Smoker	Cases	Controls
Yes	688	650
No	21	59
Total	709	709

- From Agresti (2013, 3rd ed.)

Table 3.9 Data for Fisher's Tea-Tasting Experiment

	Guess Poured First		
Poured First	Milk	Tea	Total
Milk	3	1	4
Tea	1	3	4
Total	4	4	

- From Agresti (2002, 2rd ed.)

TABLE 7.1 Primary Food Choice of Alligators

			Primary Food Choice				
Lake	Gender	Size (m)	Fish	Invertebrate	Reptile	Bird	Other
Hancock	Male	≤ 2.3	7	1	0	0	5
		> 2.3	4	0	0	1	2
	Female	≤ 2.3	16	3	2	2	3
		> 2.3	3	0	1	2	3
Oklawaha	Male	≤ 2.3	2	2	0	0	1
		> 2.3	13	7	6	0	0
	Female	≤ 2.3	3	9	1	0	2
		> 2.3	0	1	0	1	0
Trafford	Male	≤ 2.3	3	7	1	0	1
		> 2.3	8	6	6	3	5
	Female	≤ 2.3	2	4	1	1	4
		> 2.3	0	1	0	0	0
George	Male	≤ 2.3	13	10	0	2	2
		> 2.3	9	0	0	1	2
	Female	≤ 2.3	3	9	1	0	1
		> 2.3	8	1	0	0	1

- From Agresti (2002, 2rd ed.)

TABLE 7.7 Life-Length Distribution of U.S. Residents (Percent),[a] 1981

	Males		Females	
Life Length	White	Black	White	Black
0–20	2.4 (2.4)	3.6 (4.4)	1.6 (1.2)	2.7 (2.3)
20–40	3.4 (3.5)	7.5 (6.4)	1.4 (1.9)	2.9 (3.4)
40–50	3.8 (4.4)	8.3 (7.7)	2.2 (2.4)	4.4 (4.3)
50–60	17.5 (16.7)	25.0 (26.1)	9.9 (9.6)	16.3 (16.3)
Over 65	72.9 (73.0)	55.6 (55.4)	84.9 (84.9)	73.7 (73.7)

- From Agresti (2013, 3rd ed.)

Table 8.7 Outcomes for Pregnant Mice in Developmental Toxicity Study

Concentration (mg/kg per day)	Response: Nonlive	Malformation	Normal
0 (controls)	15	1	281
62.5	17	0	225
125	22	7	283
250	38	59	202
500	144	132	9

- From Agresti (2002, 2rd ed.)

TABLE 2.8 Cross-Classification of Job Satisfaction by Income

Income (dollars)	Job Satisfaction: Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied
< 15,000	1	3	10	6
15,000–25,000	2	3	10	7
25,000–40,000	1	6	14	12
> 40,000	0	1	9	11

❖ **Reading**: Agresti (2013), 1.1; Faraway (2006, 1st ed.), Preface (page v)

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.

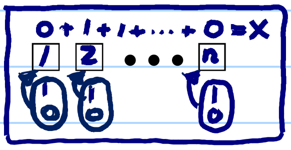
- **pmf:** $p(x) = \begin{cases} p^x(1-p)^{(1-x)}, & \text{if } x=0 \text{ or } x=1 \\ 0, & \text{otherwise} \end{cases}$
- **mgf:** $pe^t + 1 - p$
- **mean:** p
- **variance:** $p(1-p)$
- **parameter:** $p \in [0,1]$
- **example:** toss a coin once, p=probability that head occurs

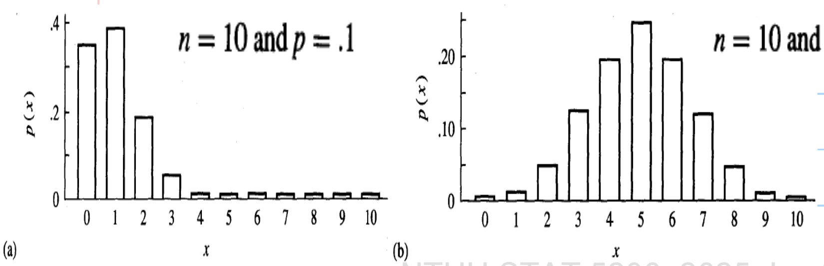
Note: If A is an event, then the indicator random variable I_A follows the Bernoulli distribution.

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.

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

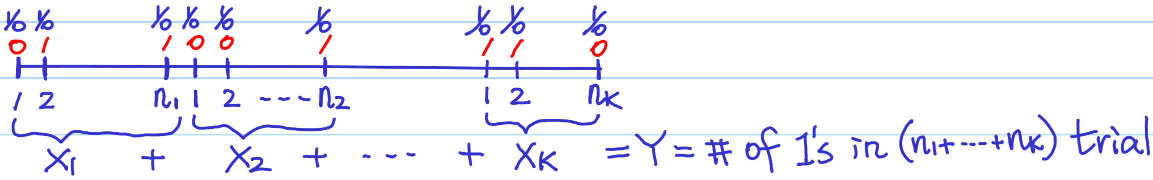




Note:
$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$.

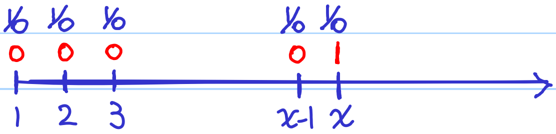
Note.

1. binomial distribution is a generalization of bernoulli distribution from 1 trial to n trials
2. Let $X_1, \ldots, X_n$ be i.i.d. $B(p)$, then $Y = X_1 + \cdots + X_n \sim B(n, p)$.
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.



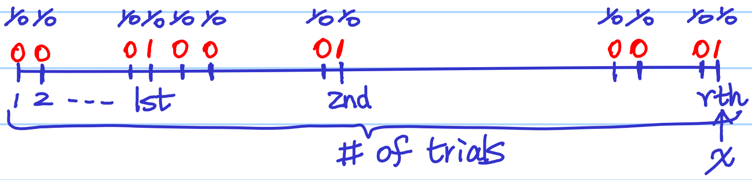
- **pmf:** $p(x) = \begin{cases} (1-p)^{(x-1)}p, & x = 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases}$
- **cdf:** $F(x) = \begin{cases} 1-(1-p)^{[x]}, & 1 \le [x] \le x < [x]+1 \\ 0, & x < 1 \end{cases}$
- **mgf:** $\frac{pe^t}{1-(1-p)e^t}$, $t < -\log(1-p)$.
- **mean:** $\frac{1}{p}$
- **variance:**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

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.

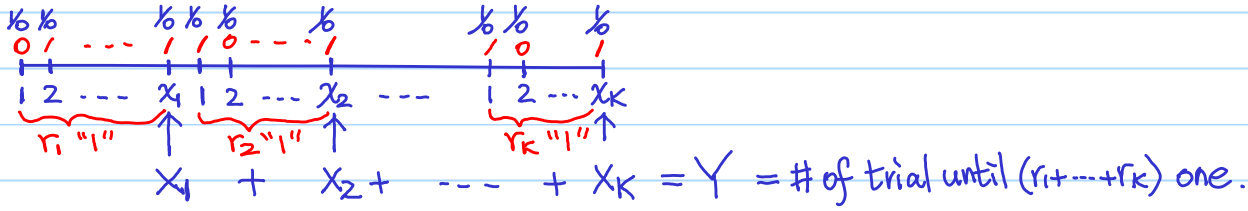


- **pmf:** $p(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{(x-r)}, & x = r, r+1, \ldots \\ 0, & \text{otherwise} \end{cases}$
- **mgf:** $\frac{p^r e^{rt}}{[1-(1-p)e^t]^r}$, $t < -\log(1-p)$.
- **mean:** $\frac{r}{p}$
- **variance:** $\frac{r(1-p)}{p^2}$
- **parameter:** $p \in [0, 1]$, $r = 1, 2, \ldots$
- **example:** lottery, # of tickets a person must purchase up to and including the rth winning ticket

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
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, $Y = X_1 + \cdots + X_k \sim NB(r_1 + \cdots + r_k, p)$.



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,

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

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:**

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

- **joint mgf:** $(p_1 e^{t_1} + \cdots + p_r e^{t_r})^n$, $t_1, \ldots, t_r \in \mathbb{R}$.
- **marginal distribution:** $X_i \sim B(n, p_i)$, $i = 1, \ldots, r$
- **mean:** $E(X_i) = np_i$, $i = 1, \ldots, n$
- **variance:** $Var(X_i) = np_i(1 - p_i)$, $i = 1, \ldots, n$
- **covariance:** $Cov(X_i, X_j) = -np_i p_j$, $i \neq j$
- **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 $(X_1, \ldots, X_r) \sim \textit{multinomial}(n, p_1, \ldots, p_r)$. 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 $q_j = \sum_{m=i_{j-1}-1}^{i_j} p_m$, $j = 1, \ldots, k$.

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

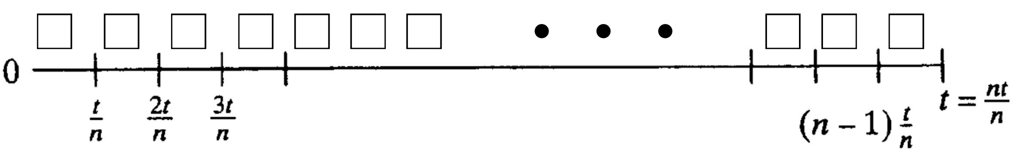
Limit of binomial distributions $X_n \sim B(n, p_n)$, where $p_n \to 0$ as $n \to \infty$ in such a way that $\lambda_n \equiv np_n \to \lambda$.

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

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

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 $\lambda = np$.

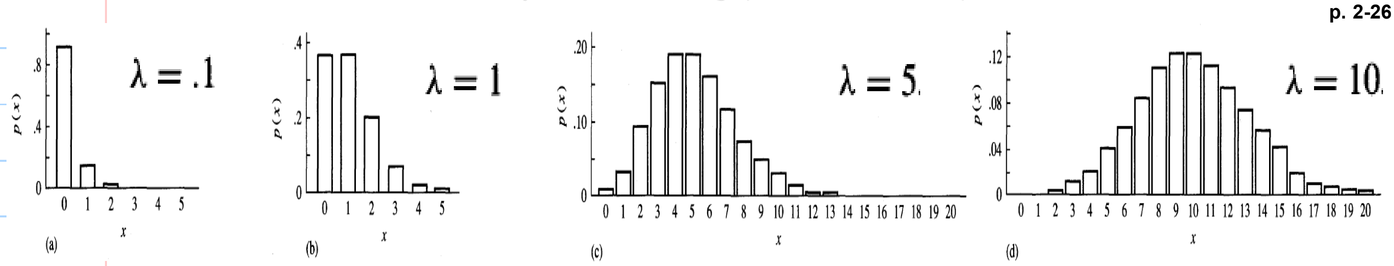


2. Let $\underline{X}$ be the <u>number of times</u> some <u>event occurs</u> in a given <u>time interval I</u>. <u>Divide</u> the interval into <u>many small subintervals I_k</u>, $k = 1, \ldots, n$, of equal length. Let $\underline{N_k}$ be the <u>number of events</u> occurring in $\underline{I_k}$.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

- **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:** $\lambda > 0$

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

- **example:** <u>number of phone calls</u> coming into an exchange during a <u>unit of time</u>



Notes.

Let $\underline{X_i} \sim \underline{P(\underline{\lambda_i})}$, $i = 1, \ldots, \underline{k}$, and $\underline{X_1, \ldots, X_k}$ are <u>independent</u>. Then,

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

2. $(\underline{X_1, \ldots, X_k} \,|\, \underline{Y = \underline{n}}) \sim \underline{\textit{multinomial}}(\underline{n}, \underline{p_1, \ldots, p_k})$, where
$$\underline{p_{\underline{i}}} = \frac{\lambda_{\underline{i}}}{\underline{\lambda_1 + \cdots + \lambda_k}} = \frac{\lambda_{\underline{i}}}{\underline{\lambda}},\ i = 1, \ldots, k.$$
The <u>converse statement</u> also <u>holds</u> with $\underline{\lambda_i} = \underline{\lambda} \times \underline{p_i}$.

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.

- **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:** $\frac{rn}{n+m}$
- **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,

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

❖ **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.