About Probability Distribution

IE241 Engineering Statistics I – Session 2

1. Random variable & probability & valid probability distribution

	Discrete (Chap 3)	Continuous (Chap 4)
Random variable	Y can assume only a finite or countably infinite number of distinct values.	Y assume uncountably infinite number of distinct values.
Probability	Defined as the sum of the probabilities of all sample points in S that are assigned the value y. Can be represented by a table or a formula that provides $p(Y = y)\ \ for\ \ all\ \ y$	Defined as the integral sum of the probability that Y falls in the interval [a,b] by $p(a \le Y \le b) = \int_a^b f(y)dy$ Note that $p(Y = y) = 0$ for $\forall y$
Probability Distribution	Can be represented in various ways. 1. (*cumulative*) distribution function 2. probability mass function (discrete) 3. probability density function (continuous) 4. moment generating function 5. probability generating function	
(*Cumulative*) distribution function $F(y) = P(Y \le y)$	Non-decreasing, [0,1]-bounded, and right-continuous 1) If $y_1 < y_2$, then $F(y_1) \le F(y_2)$ 2) $\lim_{y \to -\infty} F(y) = 0,\quad \lim_{y \to \infty} F(y) = 1$ 3) $\lim_{\substack{h \to 0 \\ h > 0}} F(y + h) = F(y+) = F(y)$	
	Always step functions	Continuous
Probability mass function (pmf) Probability density function (pdf)	$f(y) = p(Y = y)$ 1) $f(y) \ge 0,\ -\infty < y < \infty$ $f(y) = 0, \forall y \ne y_k\ \ (k = 1,2, \ldots)$ 2) $\Sigma_y f(y) = \Sigma_{k=1}^{\infty} f(y_k) = 1$ 3) $\Sigma_{a \le y \le b} f(y) = p(a \le Y \le b)$	$f(y) = \frac{dF(y)}{dy} = F'(y)$ $f(y) \ne p(Y = y) = 0$ 1) $f(y) \ge 0,\ -\infty < y < \infty$ 2) $\int_{-\infty}^{\infty} f(y)dy = 1$ 3) $\int_a^b f(y)dy = p(a \le Y \le b)$

2. Distributions with varying parameters

Please refer to the uploaded codes.

3. Relationship between Distributions

① Hypergeometric distribution vs Binomial distribution

<table>
<tr><th>Hypergeometric distribution</th><th>Binomial distribution</th></tr>
<tr><td>$Y \sim Hyper(n, D, N)$
$f(y) = P(Y = y; n, D, N) = \frac{\binom{D}{y}\binom{N-D}{n-y}}{\binom{N}{n}}$
$0 \le y \le D, 0 \le n - y \le N - D$</td><td>$Y \sim B(n, p)$
$f(y) = P(Y = y; n, p) = \binom{n}{y} p^y (1-p)^{n-y}$
$0 \le p \le 1, y = 0,1, \dots, n$</td></tr>
<tr><td>$E(Y) = \frac{nD}{N}$
$V(Y) = \frac{N-n}{N-1} n \frac{D}{N}\left(1 - \frac{D}{N}\right)$</td><td>$E(Y) = np$
$V(Y) = npq \quad (q = 1 - p)$</td></tr>
<tr><td colspan="2">

$$\text{Let } \mathrm{p} = \frac{\mathrm{D}}{\mathrm{N}} \ \ and \ \ \textbf{when } \mathbf{N} \to \infty,$$

$$E(Y) = \frac{nD}{N} = \mathrm{np}, V(Y) = \frac{N-n}{N-1} n \frac{D}{N}\left(1 - \frac{D}{N}\right) \to npq$$

What is the relationship between two distributions?

Any intuitive interpretation?

(Sketch idea)

$$\frac{\binom{D}{y}\binom{N-D}{n-y}}{\binom{N}{n}} = \frac{D!}{y!\,(D-y)!} \frac{(N-D)!}{(n-y)!\,(N-D-n+y)!} \frac{n!\,(N-n)!}{N!}$$

$$= \binom{n}{y} \frac{D(D-1)\cdots(D-y+1)}{N(N-1)\cdots(N-y+1)} \frac{(N-D)\cdots(N-D-n+y+1)}{(N-y)\cdots(N-n+1)}$$

$$\cong \binom{n}{y} \left(\frac{D}{N}\right)^y \left(1 - \frac{D}{N}\right)^{n-y}$$

</td></tr>
</table>

Thus, as $\mathbf{N} \to \infty$,
sampling without replacement (Hypergeometric) goes similar to
sampling with replacement (Binomial).

② Binomial distribution vs Poisson distribution

Binomial distribution	Poisson distribution
$Y \sim B(n,p)$ $f(y) = P(Y = y; n, p) = \binom{n}{y} p^y (1-p)^{n-y}$ $0 \le p \le 1, y = 0,1,\ldots,n$	$Y \sim Poi(\lambda)$ $f(y) = P(Y = y; \lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$ $y = 0,1,\ldots$
$E(Y) = np$ $V(Y) = npq \quad (q = 1-p)$	$E(Y) = \lambda$ $V(Y) = \lambda$
(Sketch idea) $\binom{\mathbf{n}}{\boldsymbol{y}} \mathbf{p}^{\mathbf{y}} (\mathbf{1} - \boldsymbol{p})^{\boldsymbol{n}-\boldsymbol{y}} = \frac{\boldsymbol{n}(\boldsymbol{n}-\mathbf{1}) \cdots (\boldsymbol{n}-\boldsymbol{y}+\mathbf{1}) \boldsymbol{p}^{\boldsymbol{y}} (\mathbf{1}-\boldsymbol{p})^{\boldsymbol{n}-\boldsymbol{y}}}{\boldsymbol{y}!}$ $= \frac{\frac{\mathbf{n}}{\mathbf{n}}\left(\mathbf{1}-\frac{\mathbf{1}}{\mathbf{n}}\right) \cdots \left(\mathbf{1}-\frac{\mathbf{y}-\mathbf{1}}{\mathbf{n}}\right) (\mathbf{np})^{\mathbf{y}} (\mathbf{1}-\boldsymbol{np})^{\boldsymbol{n}-\boldsymbol{y}}}{\boldsymbol{y}!}$ $\cong (\mathbf{np})^{\mathbf{y}} (\mathbf{1}-\boldsymbol{np})^{\boldsymbol{n}-\boldsymbol{y}} / \boldsymbol{y}!$ **When $\mathbf{n} \to \infty$ & $\mathbf{np_n} \to \boldsymbol{\lambda}$**	

Ex) Shin-Soo Choo, a South Korean professional baseball right fielder of Major League Baseball (MLB), has a career ratio of 50 at bats per home run (AB/HR = # of at-bats / # of home run), hitting 2 home runs in 100 at-bats on average. What is the probability that he would make 15 home runs in 400 times at-bats in 2018?

$$\#\ \text{of home runs} = \mathrm{Y} \sim \mathrm{B}(400, 0.02)$$

$$\mathrm{P}(\mathrm{Y} = 15) = \binom{400}{15} 0.02^{15} (0.98)^{385} \cong \frac{e^{-(400 \times 0.02)} (400 \times 0.02)^{15}}{15!} = 0.0090$$

③ Negative binomial distribution vs geometric distribution

Negative binomial distribution	Geometric distribution
$Y \sim Negbin(r,p)$ $f(y) = P(Y = y; r, p) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$ $0 \le p \le 1, y = r, r+1, \ldots$	$Y \sim Geo(p)$ $f(y) = P(Y = y; p) = (1-p)^{y-1} p$ $0 \le p \le 1, y = 1,2, \ldots$
$E(Y) = r/p$ $V(Y) = \frac{r(1-p)}{p^2}$	$E(Y) = \frac{1}{p}$ $V(Y) = \frac{(1-p)}{p^2}$

Let $\mathrm{W_i}$ *be the* i^{th} *success of Negative binomial distribution.* $i = 1, \ldots, r$

Then, $W_{i+1} - W_i$ *represents the # of trials after the* i^{th} *success to the next.*

$$p(\mathrm{W}_1 = x_1, W_2 - W_1 = x_2, \ldots, W_r - W_{r-1} = x_r)$$

$$= p(\mathrm{W}_1 = x_1) p(W_2 - W_1 = x_2) \ldots p(W_r - W_{r-1} = x_r)$$

$$= \{(1-\mathrm{p})^{\mathrm{x}_1 - 1}\mathrm{p}\}\{(1-\mathrm{p})^{\mathrm{x}_2 - 1}\mathrm{p}\} \cdots \{(1-\mathrm{p})^{\mathrm{x}_\mathrm{r} - 1}\mathrm{p}\}$$

$$where\ \ \mathrm{x_i} = 1,2, \cdots (\mathrm{i} = 1, \cdots \mathrm{r})$$

$$\therefore \mathrm{W}_1, W_2 - W_1, \cdots, W_r - W_{r-1} \sim Geo(p) : i.i.d.$$

$$\therefore \mathrm{W}_1 + (W_2 - W_1) + \cdots + (W_r - W_{r-1}) = \mathrm{W_r} \sim Negbin(r,p)$$

$$Y \sim Negbin(r,p) \Leftrightarrow Y \equiv Z_1 + \cdots + Z_r \ \ where\ \ Z_i \sim Geo(p), i = 1, \cdots, r$$

④ Gamma distribution vs Exponential distribution

<table>
<tr><th>Gamma distribution</th><th>Exponential distribution</th></tr>
<tr><td>$Y \sim Gamma(\alpha, \beta)$

$f(y) = P(Y = y; \alpha, \beta) = \dfrac{y^{\alpha-1} e^{-y/\beta}}{\Gamma(\alpha)\beta^{\alpha}}$

$0 \leq y < \infty$</td><td>$Y \sim Exp(\beta)$

$f(y) = P(Y = y; \beta) = \dfrac{1}{\beta} e^{-y/\beta}$

$0 \leq y < \infty$</td></tr>
<tr><td>$E(Y) = \alpha\beta$

$V(Y) = \alpha\beta^2$</td><td>$E(Y) = \beta$

$V(Y) = \beta^2$</td></tr>
<tr><td colspan="2">Let $\mathrm{W_i}$ $be\ the\ i^{th}\ success\ of\ Gamma\ distribution.\quad i = 1, \ldots, \alpha\ \ (\alpha \in \mathbb{N})$

Then, $W_{i+1} - W_i$ $represents\ the\ amount\ of\ time\ after\ the\ i^{th}\ success\ to\ the\ next.$

$p(\mathrm{W}_1 = x_1, W_2 - W_1 = x_2, \ldots, W_\alpha - W_{\alpha-1} = x_r)$

$= p(\mathrm{W}_1 = x_1) p(W_2 - W_1 = x_2) \ldots p(W_\alpha - W_{\alpha-1} = x_r)$

$where\ \ \mathrm{x_i} = 1,2, \cdots (\mathrm{i} = 1, \cdots \alpha)$

$\therefore \mathrm{W}_1, W_2 - W_1, \cdots, W_\alpha - W_{\alpha-1}\ \sim\ Exp(\beta) : i.i.d.$

$\therefore \mathrm{W}_1 + (W_2 - W_1) + \cdots + (W_\alpha - W_{\alpha-1}) = \mathrm{W}_\alpha\ \sim\ Gamma(\alpha, \beta)$

$Y \sim Gamma(\alpha, \beta) \Leftrightarrow Y \equiv Z_1 + \cdots + Z_\alpha\ \ where\ \ Z_i \sim Exp(\beta), i = 1, \cdots, \alpha$</td></tr>
</table>

Distribution	Notation	pmf / pdf	Representational definition	Moment generating function (mgf)
Bernoulli	$Y \sim Bernoulli(p)$ $0 \le p \le 1$	$p^y (1-p)^y$ $y = 0,1$	$B(1, p)$	$pe^t + q$ $q = 1 - p$
Binomial	$Y \sim B(n, p)$ $0 \le p \le 1$	$\binom{n}{y} p^y (1-p)^{n-y}$ $y = 0,1, \dots, n$	$Y \sim B(n,p) \Leftrightarrow Y \equiv Z_1 + \cdots + Z_n$ $Z_i \sim Bernoulli(p), i = 1, \cdots, n$	$(pe^t + q)^n, q = 1 - p$ $-\infty < t < \infty$
Negative binomial	$Y \sim Negbin(r, p)$ $0 < p < 1, r \in \mathbb{N}$	$\binom{y-1}{r-1} p^r (1-p)^{y-r}$ $y = r, r+1, \dots$	$Y \sim Negbin(r, p)$ $\Leftrightarrow Y \equiv Z_1 + \cdots + Z_r$ $Z_i \sim Geo(p), i = 1, \cdots, r$	$\{pe^t (1 - qe^t)^{-1}\}^r, q = 1 - p$ $t < -\log q$
Multinomial	$Y \sim Multi(n, (p_1, p_2, \dots, p_k)^t)$ $p_j \ge 0, \sum_{j=1}^{k} p_j = 1$	$\binom{n}{y_1 y_2 \cdots y_k} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k}$ $y_j = 0, \cdots, n \ \ (j = 1, \cdots k)$ $y_1 + y_2 + \cdots y_k = n$	$Y \sim Multi(n, (p_1, \dots, p_k)^t)$ $\Leftrightarrow Y \equiv Z_1 + \cdots + Z_n$ $Z_i \sim Multi(1, (p_1, \dots, p_k)^t)$ $Z_i = (Z_{i1}, \cdots Z_{ik})^t$	$(p_1 e^{t_1} + \cdots + p_k e^{t_k})^n$ $-\infty < t_j < \infty$ $j = 1, \cdots, k$
Poisson	$Y \sim Poisson(\lambda)$ $\lambda \ge 0$	$\frac{e^{-\lambda} \lambda^y}{y!}$ $y = 0,1, \cdots$		$e^{-\lambda + \lambda e^t}$ $-\infty < t < \infty$
Geometric	$Y \sim Geo(p)$ $0 < p < 1$	$(1-p)^{y-1} p$ $y = 1,2, \cdots$	$Negbin(1, p)$	$pe^t (1 - qe^t)^{-1}, q = 1 - p$ $t < -\log q$
Exponential	$Y \sim Exp(1/\lambda)$ $\lambda > 0, \lambda = \beta^{-1}$ $\lambda : rate \ \ parameter$	$\lambda e^{-\lambda y}$ $y \ge 0$	$Gamma(1, 1/\lambda)$	$\left(1 - \frac{t}{\lambda}\right)^{-1}$ $t < \lambda$
Exponential	$\boldsymbol{Y \sim Exp(\beta)}$ $\boldsymbol{\beta > 0, \beta = \lambda^{-1}}$ $\boldsymbol{\beta : scale \ \ parameter}$	$\boldsymbol{\frac{1}{\beta} e^{-\frac{y}{\beta}}}$ $\boldsymbol{y \ge 0}$	$Gamma(1, \beta)$	$(1 - \beta t)^{-1}, t < \frac{1}{\beta}$

Distribution	Notation	pmf / pdf	Representational definition	Moment generating function (mgf)
Normal	$Y \sim N(\mu, \sigma^2)$ $\mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$ $-\infty < y < \infty$	$Y \sim N(\mu, \sigma^2)$ $\Leftrightarrow Y \equiv \sigma Z + \mu, Z \sim N(0,1)$	$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ $-\infty < t < \infty$
Beta	$Y \sim Beta(\alpha_1, \alpha_2)$ $\alpha_1 > 0, \alpha_2 > 0$	$\frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y^{\alpha_1 - 1}(1-y)^{\alpha_2 - 1}$ $0 \leq y \leq 1$	$Y \sim Beta(\alpha_1, \alpha_2)$ $\Leftrightarrow Z_1/(Z_1 + Z_2)$ $Z_i \sim Gamma(\alpha_i, \beta), i = 1,2$	*Does not exist in closed form*
Gamma	$Y \sim Gamma(\alpha, 1/\lambda)$ $\alpha : shape\ \ parameter$ $\lambda : rate\ \ parameter$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y}$ $0 \leq y < \infty$	$If\ \ \alpha \in \mathbb{N}$ $Y \sim Gamma(\alpha, 1/\lambda)$ $\Leftrightarrow Y \equiv Z_1 + \cdots + Z_\alpha$ $Z_i \sim Exp(1/\lambda), i = 1, \cdots, \alpha$	$\left(1 - \frac{t}{\lambda}\right)^{-\alpha}$ $t < \lambda$
Gamma	$\boldsymbol{Y \sim Gamma(\alpha, \beta)}$ $\boldsymbol{\alpha : shape\ \ parameter}$ $\boldsymbol{\beta : scale\ \ parameter}$	$\boldsymbol{\frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-\frac{y}{\beta}}}$ $\boldsymbol{0 \leq y < \infty}$	$If\ \ \alpha \in \mathbb{N}$ $Y \sim Gamma(\alpha, \beta)$ $\Leftrightarrow Y \equiv Z_1 + \cdots + Z_\alpha$ $Z_i \sim Exp(\beta), i = 1, \cdots, \alpha$	$(1 - \beta t)^{-\alpha}$ $t < \frac{1}{\beta}$
Uniform	$Y \sim Uni(a, b)$ $a < b$	$\frac{1}{b-a}$ $a \leq y \leq b$	$Y \sim Uni(a, b)$ $\Leftrightarrow Y \equiv (b-a)Z + a, Z \sim Uni(0,1)$	$\frac{\mathrm{e}^{tb} - \mathrm{e}^{\mathrm{ta}}}{t(b-a)}$