# SI140 Discussion 11

Li Zeng, Tao Huang, Xinyi Liu

ShanghaiTech University, China
{zengli,huangtao1,liuxy10}@shanghaitech.edu.cn

## 1 Transformation

### 1.1 Transformation among Discrete R.V.s

Let $f_Y(y)$ be given. Consider function $x = g(y)$; the goal is to calculate $f_X(x)$. Let $\mathcal{X} = g(\mathcal{Y})$. For each $x_j \in \mathcal{X}$, let $\mathcal{Y}_j = \{y_{j,i}\}$ be the set of all $y \in \mathcal{Y}$ such that $g(y_{j,i}) = x_j$ (see figure, below). We claim that

$$f_x(x_j) = \sum_{y_{j,i} \in \mathcal{Y}_j} f_y(y_{j,i})$$

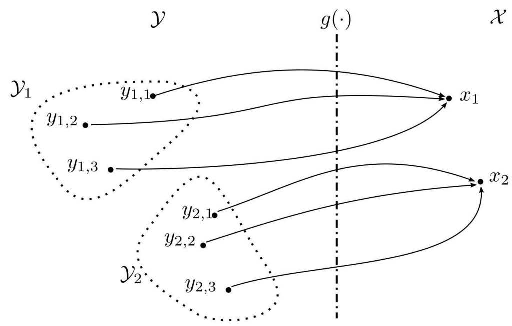


**Fig. 1.** Transformation between discrete random variables

### 1.2 Transformation among Continuous R.V.s

**One-dimension Case** Let $X$ be a continuous $r.v.$ with $PDF f_X$, and let $Y = g(X)$, where $g$ is differentiable and strictly increasing (or strictly decreasing). Then the PDF of $Y$ is given by

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

where $x = g^{-1}(y)$. The support of $Y$ is all $g(x)$ with $x$ in the support of $X$.

**Multi-dimension Case & Jacobian** Let $\mathrm{X} = (X_1, \ldots, X_n)$ be a continuous random vector with joint PDF $f_{\mathrm{X}}(x)$ and let $\mathrm{Y} = g(\mathrm{X})$ where $g$ is an invertible function from $\mathbb{R}^n$ to $\mathbb{R}^n$. Let $y = g(x)$ and suppose that all the partial derivatives $\frac{\partial x_i}{\partial y_j}$ exists and are continuous. Also assume that the determinant of the Jacobian matrix is never 0. Then the joint $PDF$ of $Y$ is

$$f_Y(y) = f_X(x) \left| det(\frac{\partial x}{\partial y}) \right| = f_X(g^{-1}(y)) \left| det(\frac{\partial g(x)}{\partial x}) \right|^{-1}$$

*Exercise 1 (Cartesian to Polar).* Let $X$ and $Y$ be i.i.d. $\mathcal{N}(0,1)$ r.v $s$, and $(R, \theta)$ be the polar coordinates for the point $(X, Y)$, so $X = R\cos\theta$ and $Y = R\sin\theta$ with $R \geq 0$ and $\theta \in [0, 2\pi)$. Find the joint PDF of $R^2$ and $\theta$. Also find the marginal distributions of $R^2$ and $\theta$, giving their names (and parameters) if they are distributions we have studied before.

## 2 Convolution

As we'll see, a convolution sum is nothing more than the law of total probability, conditioning on the value of either X or Y ; a convolution integral is analogous.

**Theorem 1 (Convolution sum and integrals).** *If $X$ and $Y$ are independent discrete, then the PMF of their sum $T = X + Y$ is*

$$\begin{aligned} P(T=t) &= \sum_x P(Y=t-x)P(X=x) \\ &= \sum_y P(X=t-y)P(Y=y). \end{aligned}$$

*If $X$ and $Y$ are independent continuous r.v.s, then the PDF of their sum $T = X + Y$ is*

$$\begin{aligned} f_T(t) &= \int_{-\infty}^{\infty} f_Y(t-x)f_X(x)dx \\ &= \int_{-\infty}^{\infty} f_X(t-y)f_Y(y)dy. \end{aligned}$$

*Exercise 2.* Let $X$ and $Y$ be *i.i.d. Expo*(1). Use a convolution integral to show that the PDF of $L = X-Y$ is $f(t) = \frac{1}{2}e^{-|t|}$ for all real $t$; this is known as the Laplace distribution.

*Solution 2.* For $t > 0$, we compute the PDF of $L$ by convolution integral:

$$\begin{aligned} f_L(t) &= \int_{-\infty}^{\infty} f_X(x+t)f_Y(x)dx \\ &= \int_{0}^{\infty} f_X(x+t)f_Y(x)dx \\ &= \int_{0}^{\infty} e^{-2x-t}dx \\ &= \frac{1}{2}e^{-t} \end{aligned}$$

Similarly, $f_L(t) = \frac{1}{2}e^t$ for $t \leq 0$. Therefore, $f(t) = \frac{1}{2}e^{-|t|}$ for all real $t$.

## 3 Order Statistics

**Definition 1 (Order Statistics).** *Given a random vector $(X_1, \ldots, X_n)$ on the probability space $(S, \mathcal{E}, P)$, for each $s \in S$, sort the components into a vector $\left(X_{(1)}(s), \ldots, X_{(n)}(s)\right)$ satisfying*

$$X_{(1)}(s) \leqslant X_{(2)}(s) \leqslant \cdots \leqslant X_{(n)}(s)$$

*The vector $\left(X_{(1)}, \ldots, X_{(n)}\right)$ is called the vector of order statistics of $(X_1, \ldots, X_n)$ Equivalently,*

$$X_{(k)} = \min\left\{\max\left\{X_j : j \in J\right\} : J \subset \{1, \ldots, n\} \& |J| = k\right\}$$

### 3.1 Marginal CDF of Order Statistics

*Note 1.* From here on, we shall assume that the original random variables $X_1, \cdots, X_n$ are i.i.d.

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with common cumulative distribution function $F$, and let $\left(X_{(1)}, \ldots, X_{(n)}\right)$ be the vector of order statistics of $X_1, \ldots, X_n$. By breaking the event $\left(X_{(k)} \leqslant x\right)$ into simple disjoint subevents, we get

$$\begin{aligned} \left(X_{(k)} \leqslant x\right) = & \bigcup \left(X_{(n)} \leqslant x\right) \\ & \bigcup \left(X_{(n)} > x, X_{(n-1)} \leqslant x\right) \\ & \bigcup \cdots \\ & \bigcup \left(X_{(n)} > x, \quad \ldots, \quad X_{(k+1)} > x, X_{(k)} \leqslant x\right) \end{aligned}$$

Each of these subevents is disjoint from the ones above it, and each has a binomial probability: $\left(X_{(n)} > x, \ldots, X_{(j+1)} > x, X_{(j)} \leqslant x\right) = (n-j$ of the random variables are $> x$ and $j$ are $\leqslant x)$ So

$$P\left(X_{(n)} > x, \ldots, X_{(j+1)} > x, X_{(j)} \leqslant x\right) = \binom{n}{j}(1-F(x))^{n-j} F(x)^j$$

Thus, the CDF of the $k^{\text{th}}$ order statistic from a sample of $n$ is:

$$F_{(k,n)}(x) = P\left(X_{(k)} \leqslant x\right) = \sum_{j=k}^{n} \binom{n}{j}(1-F(x))^{n-j} F(x)^j$$

### 3.2 Marginal PDF of Order Statistics

**Conclusion** First, we present the conclusion: the PDF of the $k^{\text{th}}$ order statistic from a sample of $n$ is

$$f_{(k,n)}(x) = n\binom{n-1}{k-1}(1-F(x))^{(n-1)-(k-1)} F(x)^{k-1} f(x)$$

**Proof 1** To understand, we construct a non-formal proof as follows (which has been emphasized in the lecture).

*Proof.* Suppose the PDF is $f$ and CDF is $F$. The density of the $k$ th order statistic is

$$\begin{aligned} & P\left(X_{(k)} \in [x, x+\epsilon]\right) = P\left(\text{one of the } X's \in [x, x+\epsilon] \text{ and exactly } k-1 \text{ of the others } < x\right) \\ = & \sum_{i=1}^{n} P\left(X_i \in [x, x+\epsilon] \text{ and exactly } k-1 \text{ of the others } < x\right) \\ = & \, nP\left(X_1 \in [x, x+\epsilon] \text{ and exactly } k-1 \text{ of the others } < x\right) \\ = & \, nP\left(X_1 \in [x, x+\epsilon]\right) P(\text{exactly } k-1 \text{ of the others } < x) \\ = & \, nP\left(X_1 \in [x, x+\epsilon]\right)\left(\binom{n-1}{k-1} P(X<x)^{k-1} P(X>x)^{n-k}\right) \end{aligned}$$

Hence

$$f_{(k)}(x) = nf(x)\binom{n-1}{k-1} F(x)^{k-1}(1-F(x))^{n-k}$$

**Proof 2** In addition, here is another formal proof for you to check.

*Proof.* Suppose $f = F'$, is the marginal PDF, then we can calculate the marginal PDF of $X_{(k)}$ by differentiating the CDF $F_{(k,n)}$, i.e.

$$\begin{aligned}
\frac{d}{dx}F_{(k,n)}(x) &= \frac{d}{dx}\sum_{j=k}^{n}\binom{n}{j}F(x)^j(1-F(x))^{n-j} \\
&= \sum_{j=k}^{n}\binom{n}{j}\frac{d}{dx}F(x)^j(1-F(x))^{n-j} \\
&= \sum_{j=k}^{n}\binom{n}{j}\left(jF(x)^{j-1}(1-F(x))^{n-j}F'(x)-(n-j)F(x)^j(1-F(x))^{n-j-1}F'(x)\right) \\
&= \sum_{j=k}^{n}\binom{n}{j}\left(jF(x)^{j-1}(1-F(x))^{n-j}-(n-j)F(x)^j(1-F(x))^{n-j-1}\right)f(x) \\
&= \sum_{j=k}^{n}\frac{n!}{(j-1)!(n-j)!}F(x)^{j-1}(1-F(x))^{n-j}f(x)-\sum_{j=k}^{n-1}\frac{n!}{j!(n-j-1)!}(1-F(x))^{n-j-1}F(x)^jf(x) \\
&= \frac{n!}{(k-1)!(n-k)!}(1-F(x))^{n-j}F(x)^{k-1}f(x) \\
&+ \sum_{j=k+1}^{n}\frac{n!}{(j-1)!(n-j)!}(1-F(x))^{n-j}F(x)^{j-1}f(x)-\sum_{j=k}^{n-1}\frac{n!}{j!(n-j-1)!}(1-F(x))^{n-j-1}F(x)^jf(t)
\end{aligned}$$

The last two terms above cancel, since using the change of variables $i = j - 1$,

$$\sum_{j=k+1}^{n}\frac{n!}{(j-1)!(n-j)!}(1-F(x))^{n-j}F(x)^{j-1} = \sum_{i=k}^{n-1}\frac{n!}{i!(n-i-1)!}(1-F(x))^{n-i}F(x)^{i}$$

So the PDF of the $k^{\text{th}}$ order statistic from a sample of $n$ is:

$$f_{(k,n)}(x) = \frac{n!}{(k-1)!(n-k)!}(1-F(x))^{n-k}F(x)^{k-1}f(x) = n\binom{n-1}{k-1}(1-F(x))^{(n-1)-(k-1)}F(x)^{k-1}f(x)$$

### 3.3 Order Statistics of Uniform

For an $Unif(0,1)$ distribution, $F(t) = t$ and $f(t) = 1$ on $[0,1]$. In this case, the PDF $f_{(k,n)}$ of the $k^{\text{th}}$ order statistic for $n$ independent $Unif(0,1)$ random variables is

$$f_{(k,n)}(t) = n\binom{n-1}{k-1}(1-t)^{n-k}t^{k-1}$$

This is an example of the Beta distribution where $r = k$ and $s = n - k + 1$.

$$X_{(k)} \sim \text{Beta}(k, n-k+1)$$

*Exercise 3.* Let $X_1, X_2$ and $X_3$ be independent $Unif(0,1)$ -distributed random variables. Prove the intuitively reasonable result that $X_{(1)}$ and $X_{(3)}$ are conditionally independent given $X_{(2)}$ and determine this (conditional) distribution.

*Solution 3.* It follows from the joint PDF of order statistics that

$$f_{X_{(1)},X_{(2)},X_{(3)}}(x_1,x_2,x_3) = 6, \quad 0 < x_1 < x_2 < x_3 < 1$$

$$f_{X_{(2)}}(x_2) = f_{\beta(2,2)}(x_2) = 6x_2(1-x_2), \quad 0 < x_2 < 1$$

which means that $f_{(X_{(1)},X_{(3)})|X_{(2)}=x}(x_1,x_3) = \frac{6}{6x(1-x)} = \frac{1}{x(1-x)}, \quad 0 < x_1 < x, \quad x < x_3 < 1$ and the statement is proven. The marginal distributions are $Unif(0,x)$ and $Unif(x,1)$.

*Exercise 4.* Let $X_1, X_2, X_3$ and $X_4$ be independent $U(0,1)$ -distributed random variables. Compute $\Pr\left(X_{(3)} + X_{(4)} \le 1\right)$

*Solution 4.* The first task is to find the joint density of $X_{(3)}$ and $X_{(4)}$. By the joint PDF of order statistics we have

$$f_{X_{(1)},\ldots,X_{(4)}}\left(x_1,\ldots,x_4\right) = 4!\prod_{k=1}^{4} f_k\left(x_k\right) = 24, \quad 0 < x_1 < x_2 < x_3 < x_4 < 1$$

which implies that

$$f_{X_{(3)},X_{(4)}}\left(x_3,x_4\right) = 24\int_0^{x_3}\int_0^{x_2} dx_1 dx_2 = 24\int_0^{x_3} x_2 dx_2 = 12x_3^2, \quad 0 < x_3 < x_4 < 1$$

From here on it is standard technique using the transformation theorem. To find the density of $X_{(3)}+X_{(4)}$ we introduce the auxiliary variable $X_{(3)}$, i.e.

$$\begin{cases} U = & X_{(3)} + X_{(4)} \\ V = & X_{(3)} \end{cases} \Longleftrightarrow \begin{cases} X_{(3)} = & V \\ X_{(4)} = & U - V \end{cases} \Longrightarrow J = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

and so it follows from the transformation theorem that

$$f_{U,V}(u,v) = f_{X_{(3)},X_{(4)}}(v, u-v) = 12v^2, \quad 0 < v < 1, \quad 2v < u < v+1$$

Hence, the marginal distribution of $U$ is given by

$$f_U(u) = \int_0^{u/2} 12v^2 dv = \frac{u^3}{2}, \quad 0 < u < 1$$
$$f_U(u) = \int_{u-1}^{u/2} 12v^2 dv = \frac{(u-2)\left(-7u^2+10u-4\right)}{2}, \quad 1 \le u < 2$$

and so we finally find that

$$\Pr\left(X_{(3)} + X_{(4)} \le 1\right) = \Pr(U \le 1) = \int_0^1 \frac{u^3}{2} du = \frac{1}{8}$$

### 3.4 Beta Distribution

**Beta Function** Recall above the PDF of Uniform order statistics, i.e.

$$f_{(k,n)}(t) = n\binom{n-1}{k-1}(1-t)^{n-k}t^{k-1}$$

Since $f_{(k,n)}$ is a density

$$\int_0^1 f_{(k,n)}(t)dt = n\binom{n-1}{k-1}\int_0^1 (1-t)^{n-k}t^{k-1} = 1$$

Or

$$\int_0^1 (1-t)^{n-k}t^{k-1} = \frac{1}{n\binom{n-1}{k-1}} = \frac{(k-1)!(n-k)!}{n!}$$

Now change variables by setting

$$r = k \quad \text{and} \quad s = n - r + 1 \text{ (so } s - 1 = n - r \text{ and } n = r + s - 1)$$

Reformulate the above equation

$$\int_0^1 (1-t)^{s-1}t^{r-1} = \frac{(r-1)!(s-1)!}{(s+r-1)!} = \frac{\Gamma(s)\Gamma(r)}{\Gamma(r+s)}$$

Recall that the Gamma function is a continuous version of the factorial, and has the property that $\Gamma(s+1) = s\Gamma(s)$ for every $s > 0$, and $\Gamma(m) = (m-1)!$ for every natural number $m$. This fact suggests the following definition of Beta function:

**Definition 2 (Beta Function).** *The Beta function is defined for $r, s > 0$ (not necessarily integers), by*

$$B(r,s) = \int_0^1 t^{r-1}(1-t)^{s-1}dt = \frac{\Gamma(s)\Gamma(r)}{\Gamma(r+s)}$$

*Remark 1.* So for integers $r$ and $s$, we see that

$$B(r+1,s+1) = \frac{\Gamma(s+1)\Gamma(r+1)}{\Gamma(r+s+2)} = \frac{(s)!(r)!}{(r+s-1)!} = (r+s)\frac{(s)!(r)!}{(r+s)!} = (r+s)\frac{1}{\binom{r+s}{r}}$$

**Beta Distribution**

**Definition 3 (Beta Distribution).** *The Beta distribution is a continuous distribution defined on the range (0,1) where the density is given by*

$$f(x) = \frac{1}{B(r,s)}x^{r-1}(1-x)^{s-1}$$

*Remark 2.* Due to the definition of Beta function, the PDF integration is ensured to be 1.

*Property 1 (Expectation).* Let $X \sim \mathrm{Beta}(r,s)$ then

$$\begin{aligned}
E(X) &= \int_0^1 x\frac{1}{B(r,s)}x^{r-1}(1-x)^{s-1}dx \\
&= \frac{1}{B(r,s)}\int_0^1, 1x^{(r+1)-1}(1-x)^{s-1}dx \\
&= \frac{B(r+1,s)}{B(r,s)} \\
&= \frac{r!(s-1)!}{(r+s)!}\frac{(r+s-1)!}{(r-1)!(s-1)!} \\
&= \frac{r!}{(r-1)!}\frac{(r+s-1)!}{(r+s)!} \\
&= \frac{r}{r+s}
\end{aligned}$$

*Property 2 (Variance).* Let $X \sim \mathrm{Beta}(r,s)$ then

$$\begin{aligned}
E\left(X^2\right) &= \int_0^1 x^2\frac{1}{B(r,s)}x^{r-1}(1-x)^{s-1}dx \\
&= \frac{B(r+2,s)}{B(r,s)} = \frac{(r+1)!(s-1)!}{(r+s+1)!}\frac{(r+s-1)!}{(r-1)!(s-1)!} \\
&= \frac{(r+1)r}{(r+s+1)(r+s)} \\
&= \frac{(r+1)r(r+s) - r^2(r+s+1)}{(r+s+1)(r+s)^2} \\
\mathrm{Var}(X) &= E\left(X^2\right) - E(X)^2 \\
&= \frac{(r+1)r}{(r+s+1)(r+s)} - \frac{r^2}{(r+s)^2} \\
&= \frac{(r+s)}{(r+s+1)(r+s)^2}
\end{aligned}$$