STA 114: Statistics

Robert L. Wolpert

1 Change of Variables

1.1 One Dimension

Let X be a real-valued random variable with pdf $f_X(x)$ and let $Y = g(X)$ for some strictly monotonically-increasing differentiable function $g(x)$; then Y will have a continuous distribution too, with some pdf $f_Y(y)$ and the expectation of any nice enough function $h(Y)$ can be computed either as

$$\begin{aligned}
\mathsf{E}[g(Y)] &= \int h\big(g(x)\big)\, f_X(x)\, dx \text{ or as} \\
&= \int h(y)\, f_Y(y)\, dy
\end{aligned}$$

Since $y = g(x)$ and $dy/dx = g'(x)$, we can write $dy = g'(x)\, dx$ and get

$$= \int h\big(g(x))\big)\, f_Y(y)\, g'(x) dx$$

so we must have

$$\begin{aligned}
f_X(x) &= f_Y(y)\, g'(x), \ \ i.e., \\
f_Y(y) &= f_X(x)/g'(x)\Big|_{x:\ y=g(x)}.
\end{aligned}$$

If g is monotonically-*decreasing* a similar formula holds with $g'(x)$ replaced by $-g'(x)$; in both cases this is:

$$= f_X(x)/|g'(x)|\Big|_{x:\ y=g(x)},$$

giving the density function for $Y = g(X)$ in terms of that for X. A similar formula holds even for non-1:1 functions $g(z)$; just sum the RHS over all x in $g^{-1}(y) = \{x : \ g(x) = y\}$ (note this is a *set*, not a number). For example, if $X \sim \mathsf{No}(0,1)$, then the pdf for $Y = g(x) = x^2$ is

$$\begin{aligned}
f_Y(y) &= \sum_{x:\ x^2=y} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \\
&= \begin{cases} 0 & y < 0, \text{ where } g^{-1}(y) = \emptyset; \\ \frac{2}{\sqrt{2\pi}} e^{-y/2}/|2\sqrt{y}| & y > 0, \text{ where } g^{-1}(y) = \pm\sqrt{y}. \end{cases} \\
&= (2\pi y)^{-1/2}\, e^{-y/2} = \frac{(1/2)^{1/2}}{\Gamma(1/2)} y^{\frac{1}{2}-1}\, e^{-y/2} \mathbf{1}_{\{y>0\}},
\end{aligned}$$

the $\mathsf{Ga}(\frac{1}{2}, \frac{1}{2})$ density function. Thus the squared Euclidean norm of a p-dimensional vector Z whose components are independent $\mathsf{No}(0,1)$ random variables would be the sum of p independent $\mathsf{Ga}(\frac{1}{2}, \frac{1}{2})$ random variables, so

$$Z'Z = \sum_{j=1}^{p} {Z_j}^2 \sim \mathsf{Ga}(p/2,\ 1/2),$$

a distribution that occurs often enough to have its own name— the "Chi squared distribution with p degrees of freedom", or χ^2_p for short.

1.2 Vectors & Matrices

A *vector* $x \in \mathbb{R}^p$ is an ordered sequence of p real numbers, its "coordinates." We usually won't use any special notation (like $\mathbf{x}$ or $\vec{x}$) to distinguish vectors from other variables; the context should make it clear (after some practice!). An $r \times c$ *matrixc* A is a rectangular array of r rows and c columns whose entries are denoted by a_{ij} (the ith row, jth column) for $1 \le i \le r$, $1 \le j \le c$. We often view vectors as *one-column* matrices, so

$$x = (x_1, \cdots, x_p)' = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}.$$

Many (but not all) matrices in statistics are *square*. Any square $p \times p$ matrix has a "determinant" $\det(A)$, with the properties:

$$\det(c\,A) = c^p \det(A) \quad \det(A') = \det(A) \quad \det(AB) = \det(A)\det(B) \quad (1)$$

1.3 Random Vectors

Similarly if X is a *vector*-valued random variable taking values in $\mathbb{R}^p$, with (joint) pdf $f_X(x) \geq 0$ defined on $\mathbb{R}^p$, and if $g : \mathbb{R}^P \to \mathbb{R}^p$ is a 1:1 differentiable function, then $Y = g(X)$ also has a density function, then instead of "$dy/dx = g'(x)$" we have a $p \times p$ matrix

$$J(x) = \left\{ \frac{\partial y_i}{\partial x_j} \right\}$$

of partial derivatives, called the "Jacobian," and change of variables takes the form

$$\begin{aligned} \mathsf{E}[g(Y)] &= \int h\big(g(x)\big)\, f_X(x)\, dx \\ &= \int h(y)\, f_Y(y)\, dy \\ &= \int h\big(g(x)\big))\, f_Y(y)\, |\det J(x)| dx \end{aligned}$$

so we must have

$$f_Y(y) = f_X(x)/|\det J(x)|, \qquad x \in g^{-1}(y) \tag{2}$$

2 Examples

2.1 Multivariate Normal

Let A be an invertible $p \times p$ matrix and μ a vector (which we view as a $p \times 1$ matrix), and let $Z = (Z_1, \cdots, Z_p)'$ be a p-dimensional vector of independent standard normal random variables $\{Z_j\} \overset{\text{iid}}{\sim} \mathsf{No}(0,1)$, with joint pdf $f_Z(z) = (2\pi)^{-p/2} \exp(-z'z/2)$. Then

$$X = \mu + A\,Z$$

is a p-dimensional normal vector with mean μ and covariance matrix

$$\begin{aligned} C &= \mathsf{E}(X - \mu)(X - \mu)' \\ &= \mathsf{E}(A\,Z)(A\,Z)' \\ &= \mathsf{E}\ [A\,Z\,Z'\,A'] \quad \text{(because } \mathsf{E}[Z\,Z'] = I_p) \\ &= A\,A' \end{aligned}$$

so by Eqn. (2), $X = g(Z) = \mu + AZ$ (with Jacobian $J(z) = \partial g_i(z)/\partial z_j = A_{ij}$ and inverse $g^{-1}(x) = A^{-1}(X - \mu)$) has pdf:

$$
\begin{aligned}
f_X(x) &= f_Z(z)/|\det J(z)| && \text{(3a)}\\
&= (2\pi)^{-p/2}\exp\Big[-(X-\mu)'(A^{-1})'(A^{-1})(X-\mu)/2\Big]/|\det A| && \text{(3b)}\\
&= (2\pi)^{-p/2}\exp\Big[-(X-\mu)(A\,A')^{-1}(X-\mu)/2\Big]/\sqrt{\det A\,A'} && \text{(3c)}\\
&= \frac{1}{\sqrt{\det 2\pi C}}\,e^{-(X-\mu)'C^{-1}(X-\mu)/2}, && \text{(3d)}
\end{aligned}
$$

the pdf for $X \sim \mathsf{No}(\mu, C)$. Eqn. (3a) is just the multivariate CoV of Eqn. (2); Eqn. (3b) is from instantiating $f_Z(z)$, $g^{-1}(x)$, and $J(z)$ (using the elementary linear algebra fact that $(AB)' = B'A'$ for any two $p \times p$ matrices A, B); Eqn. (3c) uses the elementary linear algebra facts that $(AB)^{-1} = B^{-1}A^{-1}$ and $\det(AB) = (\det A)(\det B)$ for any two $p \times p$ matrices A, B); and Eqn. (3d) uses the facts that $C = AA'$ and that $\det(cA) = c^p \det A$.

The special case of $C = I_p$ and $\mu = 0$ reduces to the joint pdf of p iid $\mathsf{No}(0, 1)$ random variables, while the special case of $p = 1$ (with $C = \sigma^2 \in \mathbb{R}_1$) is the familiar $\mathsf{No}(\mu, \sigma^2)$ density.

2.2 Another CoV example: Gamma, Beta

Let $X \sim \mathsf{Ga}(\alpha, \lambda)$ and $Y \sim \mathsf{Ga}(\beta, \lambda)$ be independent for some $\alpha, \beta, \lambda > 0$, and set $U = X + Y$, $V = X/(X + Y)$. We can think of (U, V) as the two components of the function $g : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$
\begin{aligned}
g(x,y) &= [(x+y), x/(x+y)]'\\
\frac{\partial}{\partial x} g(x,y) &= [1, y/(x+y)^2]'\\
\frac{\partial}{\partial y} g(x,y) &= [1, -x/(x+y)^2]'\\
J(x,y) &= \begin{bmatrix} 1 & 1 \\ y/(x+y)^2 & -x/(x+y)^2 \end{bmatrix}\\
\det J &= -1/(x+y)\\
g^{-1}(u,v) &= [uv, u(1-v)]'
\end{aligned}
$$

The joint pdf for X, Y is:

$$
f_{XY}(x,y) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}\, y^{\beta-1}\, e^{-\lambda(x+y)}, \qquad x, y > 0
$$

so that of U, V, by CoV, is:

$$\begin{aligned}
f_{UV}(u,v) &= f_{XY}(x,y)/|J(x,y)| \\
&= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)}(uv)^{\alpha-1}\,(u(1-v))^{\beta-1}\,e^{-\lambda u} \times u, \qquad 0<u<\infty,\ 0<v<1 \\
&= \left\{\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)}u^{\alpha+\beta-1}e^{-\lambda u}\mathbf{1}_{\{0<u<\infty\}}\right\}\left\{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}v^{\alpha-1}(1-v)^{\beta-1}\mathbf{1}_{\{0<v<1\}}\right\}
\end{aligned}$$

so U, V are independent with the $\mathsf{Ga}(\alpha+\beta,\lambda)$ and $\mathsf{Be}(\alpha,\beta)$ distributions, respectively.