Application of the Change of Variable Theorem:

Generate the Dirichlet-distribution →
general extension of the Beta distribution.

(1) Start with independent random variables $X_1, X_2, \ldots, X_n$

$X_i \sim \Gamma(\alpha_i, \lambda)$.

(2) Define transformations:

$$U_1 = \frac{X_1}{X_1 + X_2 + \cdots + X_n} = g_1(X_1, \ldots, X_n)$$

$$U_2 = \frac{X_2}{X_1 + \cdots + X_n} = g_2(X_1, \ldots, X_n)$$

$$\vdots$$

$$U_{n-1} = \frac{X_{n-1}}{X_1 + \cdots + X_n} = g_{n-1}(X_1, \ldots, X_n)$$

$$U_n = X_1 + X_2 + \cdots + X_n = g_n(X_1, \ldots, X_n)$$

$X_1 = U_1 U_n$

$X_{n-1} = U_{n-1} U_n$

$X_n = U_n(1 - U_1 - \cdots - U_{n-1})$

$$\begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix} = g\left(\underline{X}\right). \qquad \underline{X} = (X_1, \dots, X_n)$$

$$= \begin{pmatrix} g_1(\underline{X}) \\ \vdots \\ g_n(\underline{X}) \end{pmatrix}$$

Goal: Find the joint density of

$(U_1, \dots, U_n)$.

First write down the joint density

of $(X_1, \dots, X_n)$.

$$f_{\underline{X}}(x_1, \dots, x_n) = \prod_{j=1}^{n} \frac{\lambda^{\alpha_j}}{\Gamma(\alpha_j)} e^{-\lambda x_j} x_j^{\alpha_j - 1} \mathbb{1}(x_j > 0)$$

$$= \frac{\lambda^{\sum \alpha_j}}{\prod \Gamma(\alpha_j)} e^{-\lambda \sum x_j} \prod_{j=1}^{n} x_j^{\alpha_j - 1} \mathbb{1}\{x_1 > 0, \dots, x_n > 0\}$$

Need to find the inverse transformation to g, i.e. express $(x_1, \ldots, x_n)$ in terms of $(u_1, \ldots, u_n)$

$$x_1 = h_1(u_1, \ldots, u_n) = u_1 u_n$$

$$x_2 = h_2(u_1, \ldots, u_n) = u_2 u_n$$

$$\vdots$$

$$x_{n-1} = h_{n-1}(u_1, \ldots, u_n) = u_{n-1} u_n$$

$$x_n = h_n(u_1, \ldots, u_n) = u_n \left[1 - \sum_{j=1}^{n-1} u_j\right]$$

$$\boxed{h(u_1, \ldots, u_n) = \begin{pmatrix} h_1(u_1, \ldots, u_n) \\ \vdots \\ h_n(u_1, \ldots, u_n) \end{pmatrix}}$$

$(u_1, \ldots, u_n) \in S_{n-1} \times (0, \infty)$ w.p. 1

$S_{n-1} = \{(w_1, \ldots, w_{n-1}) : w_i \geq 0,\ 0 < \sum w_i < 1\}$

$$\left\{ u_1 + u_2 + \cdots + u_{n-1} = \frac{x_1 + \cdots + x_{n-1}}{x_1 + \cdots + x_n} \right\} \in (0,1)$$

$$f_{(U_1, \ldots, U_n)}(u_1, \ldots, u_n) = f_X(h(u_1, \ldots, u_n)) \, J_h(u_1, \ldots, u_n)$$

$$h_i(u_1, \ldots, u_n) = u_i u_n, \quad 1 \le i \le n-1$$

$$h_n(u_1, \ldots, u_n) = u_n (1 - u_1 - \cdots - u_{n-1})$$

$$\nabla_h (u_1, \ldots, u_n)$$

$h_2(u_1, \ldots, u_n) = u_2 u_n$

$$= \begin{bmatrix} \frac{\partial h_1}{\partial u_1} = u_n, & 0, & 0, & \cdots & u_1 \\ \frac{\partial h_2}{\partial u_1} = 0, & u_n & \cdots & & u_2 \\ \vdots & & \ddots & & \\ \frac{\partial h_{n-1}}{\partial u_1} = 0, & \cdots & & u_n & u_{n-1} \\ -u_n & -u_n & \cdots & -u_n & 1 - \sum_{i=1}^{n-1} u_i \end{bmatrix}$$

diag

$$\operatorname{abs} \left| \nabla_h (u_1, \ldots, u_n) \right| = J_h(u_1, \ldots, u_n)$$

$$= u_n^{\,n-1}$$

$$f_{\underline{u}}(u_1, \dots, u_n)$$

$$= f_{\underline{x}}(h(\underline{u}))\, u_n^{n-1}\, 1\{0 < \sum_{i=1}^{n-1} u_i < 1,\ u_n > 0\}$$

$$f_{\underline{x}}(x_1, \dots, x_n) = \frac{\lambda^{\sum \alpha_j}}{\prod \Gamma(\alpha_j)} e^{-\lambda(\sum_{j=1}^{n} x_j)} \prod_{j=1}^{n} x_j^{\alpha_j - 1}$$

u_n

$$1\{x_1, \dots, x_n > 0\}$$

$$f_{\underline{x}}(h(u))\, u_n^{n-1}$$

{ add up powers of u_n in bottom-most line to get $u_n^{(\sum_1^n \alpha_j - 1)}$

$$= \frac{\lambda^{\sum \alpha_j}}{\prod \Gamma(\alpha_j)} e^{-\lambda u_n} \prod_{j=1}^{n-1} (u_j u_n)^{\alpha_j - 1} (u_n(1 - u_1 - \dots - u_{n-1}))^{\alpha_n - 1} \times u_n^{n-1}$$

$$= \frac{\lambda^{\sum \alpha_j}}{\prod \Gamma(\alpha_j)} e^{-\lambda u_n} \prod_{j=1}^{n-1} u_j^{\alpha_j - 1} (1 - u_1 - \dots - u_{n-1})^{\alpha_n - 1}$$

$$\left\{ \left(\prod_{j=1}^{n-1} u_n^{\alpha_j - 1} \right) u_n^{\alpha_n - 1}\, u_n^{n-1} \right.$$

Rearranging terms:

$$f_{\underline{U}}(u_1, \dots, u_n)$$

$$= \frac{\lambda^{\sum \alpha_j}}{\prod \Gamma(\alpha_j)} e^{-\lambda u_n} u_n^{\sum_{j=1}^{n} \alpha_j - 1} \, \mathbb{1}(u_n > 0) \quad \Psi_1$$

$$\times \left(u_1^{\alpha_1 - 1} \cdots u_{n-1}^{\alpha_{n-1}-1} (1 - u_1 - \dots - u_{n-1})^{\alpha_n - 1} \, \mathbb{1}\left(0 < \sum_1^{n-1} u_j < 1\right) \right)$$

$U_n \sim$? By the additive property for Gamma random variables,

$\underline{U_n} \sim \Gamma\left(\sum_{j=1}^{n} \alpha_j, \lambda\right)$, for example m.g.f's.

$$= \underline{\Psi_1(u_n)} \times \underline{\Psi_2(u_1, \dots, u_{n-1})}$$

$$\Rightarrow U_n \perp\!\!\!\perp (U_1, \dots, U_{n-1})$$

independent of

Marginal density of U_n has to be a multiple of ψ_1.

$$\underline{\psi_1}(u_n) = \frac{\Gamma\left(\sum_{j=1}^{n}\alpha_j\right)}{\prod_{j=1}^{n}\Gamma(\alpha_j)} \underline{\text{Gamma}\left(\sum\alpha_j, \lambda\right)} \underline{\text{density}\,(u_n)}$$

marginal density of U_n is $\text{Gamma}(\sum\alpha_j, \lambda)$

$$f_{\underline{U}}(u_1, \ldots, u_n)$$

$$= \frac{\Gamma\left(\sum_{j=1}^{n}\alpha_j\right)}{\prod_{j=1}^{n}\Gamma(\alpha_j)} \boxed{\underline{\text{Gamma}}_{\sum_1^n \alpha_j, \lambda}(u_n)} \times \psi_2(u_1, \ldots, u_{n-1})$$

So:

$$\underline{f_{U_1,\ldots,U_{n-1}}(u_1, \ldots, u_{n-1})} = \frac{\Gamma\left(\sum_{j=1}^{n}\alpha_j\right)}{\prod \Gamma(\alpha_j)} \prod_{j=1}^{n-1} u_j^{\alpha_j - 1} \times (1-u_1-\cdots-u_{n-1})^{\alpha_{n-1}}$$

$$\times \mathbb{1}\left(0 < \sum_{i=1}^{n} u_i < 1\right)$$

Precisely $\text{Beta}(\alpha_1, \alpha_2)$ when $n=2$.

This is the Dirichlet distribution with parameters $(\alpha_1, \alpha_2, \ldots, \alpha_n)$.

Linear Prediction and Bivariate Normal Random Variables:

From Page 49 of Statistics 510

Notes:

Given a pair of random variables (X, Y) with some joint distribution.

Goal: Predict Y using some (measurable) function of X, say $\phi(X)$.

What is the best predictor of Y using some $\phi(X)$?

Criterion: Mean Squared Error Criterion

look at $\mathbb{E}[(Y - \phi(X))^2]$

Integrated measure of discrepancy over the dist. of (X, Y).

Consider:

$$\mathbb{E}\left[\left(y - \phi(x)\right)^2\right]$$

$$= \mathbb{E}\left[\left(y - \mathbb{E}(y|x) + \mathbb{E}(y|x) - \phi(x)\right)^2\right]$$

$$= \mathbb{E}\left[\left(y - \mathbb{E}(y|x)\right)^2\right] \rightarrow T1 \rightarrow \text{not relevant to optimization}$$

$$+ 2\,\mathbb{E}\left[\left(y - \mathbb{E}(y|x)\right)\left(\mathbb{E}(y|x) - \phi(x)\right)\right] \rightarrow T2$$

$$+ \mathbb{E}\left[\left(\mathbb{E}(y|x) - \phi(x)\right)^2\right]$$

$$T1 = \mathbb{E}\left[\mathbb{E}\left(\left(y - \mathbb{E}(y|x)\right)^2 \middle| x\right)\right]$$

$$= \mathbb{E}\left[\mathrm{Var}(y|x)\right]$$

$$T2 = 0$$

T2

$$= 2\,\mathbb{E}\Big[\underbrace{\big(y - \mathbb{E}(y|x)\big)\big(\mathbb{E}(y|x) - \phi(x)\big)}_{G(x,y)}\Big]$$

looking at: $\mathbb{E}\,G(x,y)$

$$= \mathbb{E}\left[\mathbb{E}\big(G(x,y)\,|\,x\big)\right]$$

$$\mathbb{E}\Big[\big\{\big(y - \mathbb{E}(y|x)\big)\big(\mathbb{E}(y|x) - \phi(x)\big)\big\}\,\Big|\,x\Big]$$

$$\boxed{\begin{aligned}&\mathbb{E}\left[\mathbb{E}(y|x)\,|\,x\right]\\ &= \mathbb{E}(y|x).\end{aligned}}$$

$$= \big(\mathbb{E}(y|x) - \phi(x)\big) \times \underbrace{\mathbb{E}\left[\big(y - \mathbb{E}(y|x)\big)\,\big|\,x\right]}_{\mathbb{E}[y|x] \;-\; \mathbb{E}\left[\left(\mathbb{E}(y|x)\right)\,|\,x\right]}$$

$$= 0$$

Then:

$$\mathbb{E}\left[(Y - \phi(X))^2\right]$$

$$= \underline{\mathbb{E}\left[\mathrm{Var}(Y|X)\right]} + \underline{\mathbb{E}\left[(\mathbb{E}(Y|X) - \phi(X))^2\right]}$$

minimize

$$\mathbb{E}\left[(\mathbb{E}(Y|X) - \underline{\phi(X)})^2\right]$$

So choosing $\phi(X) = \mathbb{E}(Y|X)$ makes the term 0, so:

$$\boxed{\phi_{opt}(X) = \mathbb{E}(Y|X)}$$

A LOT of statistics deals with estimating with $\mathbb{E}(Y|X)$ using MODELS. (Regression function: $\mathbb{E}(Y|X)$)

Basic Model is taking $\phi(X) = \alpha + \beta X$

Linear Regression Model: Estimate Y using $\phi_{linear}(x) = \alpha + \beta X$ and find the best among all linear functions

i.e find

$$\underline{(\alpha_0, \beta_0)} = \underset{(\alpha, \beta)}{\operatorname{argmin}} \mathbb{E}\left[(Y - \alpha - \beta X)^2\right]$$

$\phi_{linear}(x)$ is monotone $\uparrow$ or $\downarrow$ in X, so may actually be a BAD model.

Good Model: say Y = Weight, X = Height

$\downarrow$

Linear Regression

because $H \uparrow \Rightarrow W \uparrow$

the true $\phi(x)$ i.e $\phi_{opt}(X) = \mathbb{E}(Y|X)$ may still not be linear.

$\phi_{\text{linear}}(x)$ is a BAD model, say when $\mathbb{E}(Y|X) = \underbrace{x^2 - x}$, and $x \in [0, 1]$.

<u>Example</u>: $Y = X^2 - X + \varepsilon$

$$\left.\begin{aligned} &X \sim \text{Unif}(0,1) \\ &\varepsilon \sim N(0, 1/10), \quad \varepsilon \perp\!\!\!\perp X. \end{aligned}\right\}$$

Check: $\mathbb{E}[Y|X] = X^2 - X$

since $\mathbb{E}(\varepsilon|X) = 0$.

Goal: Find the Best Linear Predictor of Y in terms of X.

Minimize: $\mathbb{E}\left[(Y - \alpha - \beta X)^2\right]$

over (α, β).