

3-2b: continuous distributions

Showed that $\sqrt{n} = \sqrt{E} \int_0^{\infty} e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2}} \int_0^{\infty} e^{-\frac{1}{2}x^2} dx$ Prove/Show

Also obtain $\Gamma(1/2) = \sqrt{\pi}$

Recall: Let $X \sim N(\mu, \sigma^2)$, then $Y = X^2 \sim \chi^2_{df=1}$

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$$

$$= \frac{1}{\sqrt{\pi}} \frac{1}{y^{1/2}} e^{-y/2} \text{ since } \Gamma(1/2) = \sqrt{\pi}$$

which is pdf for $\text{Gamma}(a=1/2, \beta=2)$

Recall (Poisson Approx): $X \sim \text{Binomial}(n, p)$, then $\begin{cases} E(X) = np \\ V(X) = np(1-p) \end{cases}$

Let $\sum_{i=1}^n p_i \rightarrow 0$, then $X \xrightarrow{d} \text{Poisson}(n)$.

Normal Approx: when $\begin{cases} np \rightarrow \infty \\ np(1-p) \rightarrow \infty \end{cases}$

then $X \approx \text{Normal}(np, np(1-p))$

Remark: we will later show via CLT that

$$\frac{1}{\sqrt{n}}(X - np) \xrightarrow{d} \text{Normal}(0, p(1-p)) \text{ if } \sum_{i=1}^n p_i \rightarrow p_0$$

$$\Rightarrow \frac{1}{\sqrt{n}}(X - np) \approx \text{Normal}(0, p(1-p))$$

$$X - np \approx \text{Normal}(0, np(1-p))$$

$$X \approx \text{Normal}(np, np(1-p))$$

Ex: Let $X \sim \text{Binomial}(n=15, p=0.6)$

$$\text{then } E(X) = np = 15 \quad \text{and } V(X) = np(1-p) = 2.4 \text{ (is 7)} \quad \text{Poisson approx. not appropriate; wrong!}$$

Now use Normal approx:

$$P(X \leq 15) \approx P(Y \leq 15), Y \sim \text{Normal}(15, 2.4)$$

$$= P\left(Z \leq \frac{15-15}{2.4}\right) = 0.5$$

vs. direct computation, i.e.

$$P(X \leq 15) = \sum_{x=0}^{15} \binom{15}{x} 0.6^x (0.4)^{15-x} = 0.267$$

Another way, use the continuity correction:

$$\begin{aligned} P(X \leq 15), X \in \mathbb{N} \\ \approx P(Y \leq 13.5), Y \in \mathbb{R} \\ = P\left(Z \leq \frac{13.5-15}{2.4}\right) = 0.271 \end{aligned}$$



Without this correction, we tend to underestimate prob's at the tails.

3-2c: continuous distributions

Def(Beta Dist.):= Let $X \sim \text{Beta}(\alpha, \beta)$, $\alpha, \beta > 0$

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \text{ for } x \in (0, 1)$$

Remarks: Beta Function: $B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1}$

$$\text{Fact: } B(\alpha, \beta) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha+\beta)} \Rightarrow \text{Beta}(n, m) = \frac{\Gamma(n) \Gamma(m)}{\Gamma(n+m)} x^{n-1} (1-x)^{m-1}$$

Beta pdf's for different (α, β) :

$$\begin{cases} \alpha=\beta=1 \Rightarrow \text{Beta}(1,1) \equiv \text{Uniform}(0,1) \\ \alpha>1, \beta>1 \Rightarrow \text{Beta is unimodal} \\ \alpha<\beta \Rightarrow \text{Beta is } \text{U} \text{ or } \text{M} \\ \alpha>\beta \Rightarrow \text{Beta is } \text{M} \text{ or } \text{U} \end{cases}$$

$$\begin{aligned}
 \text{MOMENTS: } E(X^k) &= \frac{1}{B(\alpha, \beta)} \int_0^1 x^k x^{\alpha-1} (1-x)^{\beta-1} dx \\
 &= \frac{1}{B(\alpha, \beta)} B(\alpha+k, \beta) \\
 &= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+k)/\Gamma(\beta)}{\Gamma(\alpha+k+\beta)} \\
 &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta+k+\alpha)} \\
 \text{DEF: } E(X) &= \frac{\Gamma(\alpha+\beta)/\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(\beta+\alpha+1)}
 \end{aligned}$$

$$\begin{aligned}
 \text{where } \Gamma(\alpha+k) &= \alpha \cdot \Gamma(\alpha) \\
 &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta+\alpha)} = \frac{\alpha}{\alpha+\beta}
 \end{aligned}$$

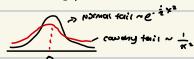
$$\text{Similarly, } E(X^2) = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+2)}{\Gamma(\alpha)\Gamma(\beta+\alpha+2)} = \frac{\alpha(\alpha+1)}{\alpha+\beta+1}$$

$$\begin{aligned}
 \text{so, } \text{Var}(X) &= E(X^2) - E(X)^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
 &= \frac{\alpha}{(\alpha+\beta)} \frac{(\alpha+1)(\alpha+\beta) - \alpha(\alpha+2)}{(\alpha+\beta)(\alpha+\beta+1)} \\
 &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
 \end{aligned}$$

DEF (Cauchy Dist): Let $X \sim \text{Cauchy}(\theta)$. $\theta \in \mathbb{R}$.

$$f(x) = \frac{1}{\pi(1+x^2)}, x \in \mathbb{R}$$

REMARKS: Cauchy pdf also bell-shaped:



θ is the median (location) of X

RECALL: $\{E(X) = \theta\}$

PROP: Let $X, Y \stackrel{iid}{\sim} N(\mu, \sigma^2)$. Then $\frac{X}{Y} \sim \text{Cauchy}(\mu)$

DEF (Log-Normal Dist): If $Y \sim N(\mu, \sigma^2)$, then $X = e^Y \sim \text{Log-Normal}, i.e.$

$$Y \sim \text{log}(X) \sim N(\mu, \sigma^2) \text{ s.t.}$$

$$f_{X|Y}(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} \exp\left(-\frac{1}{2\sigma^2} (\log(x) - \mu)^2\right)$$

3-3: Exponential Families

DEF: A family of pdf or pmf is called an exponential family if it has the form

$$\begin{aligned}
 f(x|\theta) &= c(\theta) h(x) \exp \left\{ \sum_{i=1}^k w_i(\theta) t_i(x) \right\}, x \in \mathbb{R} \\
 &= c(\theta) h(x) \exp \langle w(\theta), t(x) \rangle / c(\theta)
 \end{aligned}$$

REMARKS: where $w(\theta) = (w_1(\theta), \dots, w_k(\theta))$
 $t(x) = (t_1(x), \dots, t_k(x))$

\hookrightarrow vector of sufficient statistics (coming later)

Here,
 $h(x) \geq 0$
 $t(x) = (t_1(x), \dots, t_k(x))$ depends only on θ (data)
 $w(\theta) = (w_1(\theta), \dots, w_k(\theta))$ depends only on θ (parameters)

clearly, $c(\theta)$ is the reciprocal of the normalizing constants, i.e.

$$\int f(x|\theta) = 1 \Rightarrow \frac{1}{c(\theta)} = \int h(x) e^{\langle w(\theta), t(x) \rangle} dx \quad (\text{or sum if discrete})$$

EX: Binomial: Let $X \sim \text{Binomial}(n, p)$.

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \dots, n$$

$$= (1-p)^n \left(\frac{p}{1-p}\right)^x$$

$$\text{where } e^x = \left(\frac{p}{1-p}\right)^x$$

$$= (1-p)^n \binom{n}{x} \exp \left\{ x \log \frac{p}{1-p} \right\}$$

$$\uparrow \quad \uparrow \quad \uparrow \quad \uparrow \quad \uparrow$$

$$c(\theta) \quad h(x) \quad t_i(x) \quad w_i(\theta) \quad k = 1$$

Ex(Normal): Let $X \sim N(\mu, \sigma^2)$.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \quad x \in \mathbb{R}$$

$$\theta = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2 + \frac{2\mu}{\sigma^2}x + \frac{\mu^2}{\sigma^2}}$$

\uparrow \uparrow \uparrow
w.r.t. μ w.r.t. σ w.r.t. x

Remarks: other examples of dist's which are in the exponential family:

Poisson, Geometric, Negative, Gamma, Exponential, Lognormal

Among dist's not in this exp. family are:

Cauchy, Rithms

Thm: If X is a RV with pdf/mf in the exponential family of form (48) then,

$$(48): E\left[\frac{\partial}{\partial \theta_i} \frac{\partial \ln L(\theta)}{\partial \theta_j} t_i(x)\right] = \frac{\partial}{\partial \theta_j} \log(L(\theta)) \text{, } i = 1, \dots, k$$

$$\text{Var}\left(\frac{\partial}{\partial \theta_i} \frac{\partial \ln L(\theta)}{\partial \theta_j} t_i(x)\right) = \frac{\partial^2}{\partial \theta_j^2} \log(L(\theta)) -$$

$\left. \begin{array}{c} \uparrow \\ - E\left(\frac{\partial}{\partial \theta_i} \frac{\partial \ln L(\theta)}{\partial \theta_j} t_i(x)\right) \end{array} \right\}$ **sum of**

Remark: In words: differentiating the logarithm of the normalizing constant w.r.t. a parameter results in suitable expectations of sufficient statistics

Corollary (Simplex): Suppose

$$\theta(\omega) = \theta = (\theta_1, \theta_2, \dots, \theta_n) \in \mathbb{R}^n$$

$$t(x) = (t_1(x), t_2(x), \dots, t_n(x)) \in \mathbb{R}^n$$

then $f(x|\theta) = c(\theta) \ln(\theta) \exp(-\theta_1 t_1(x) - \dots - \theta_n t_n(x))$ and

$$(48): E(t_i(x)) = \frac{\partial}{\partial \theta_i} \log(L(\theta))$$

$$(49): \text{Var}(t_i(x)) = \frac{\partial^2}{\partial \theta_i^2} \log(L(\theta))$$

Proof: proceeds simply via calculus

$$\text{Ex(Binomial): } f_X(k|p) = (1-p)^n \binom{n}{k} \exp\left\{k \log \frac{p}{1-p}\right\}$$

\uparrow \uparrow \uparrow
c(\theta) $\ln(\theta)$ $t_i(x)$ $\ln(t_i(x))$, $k=1$

$$c(\theta) = (1-p)^n, \quad \theta = p$$

$$\frac{\partial}{\partial \theta} \log(L(\theta)) = \frac{\partial}{\partial p} n \ln \log(1-p) = \frac{n}{1-p}$$

$$\frac{\partial^2}{\partial \theta^2} \ln(L(\theta)) = \frac{\partial}{\partial p} (1-p \ln(1-p)) \quad ?$$

$$= \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}$$

Applying the thm:

$$\frac{\partial}{\partial \theta} \log(L(\theta)) = E\left(\frac{\partial}{\partial \theta} \ln(L(\theta)) t_i(x)\right)$$

$$\frac{n}{1-p} = E\left(\frac{1}{p(1-p)} x\right)$$

$$n = E\left(\frac{x}{p}\right)$$

$$\Rightarrow E(x) = np$$

$$\text{Ex(Normal): Given } f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(-\frac{1}{2\sigma^2}x^2 + \frac{2\mu}{\sigma^2}x + \frac{\mu^2}{\sigma^2}\right)$$

\uparrow \uparrow \uparrow \uparrow
c(\theta) $\ln(\theta)$ $t_i(x)$ $\ln(t_i(x))$

$$w(\theta) = \left(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2}\right)$$

$$\frac{\partial}{\partial \theta} \log(L(\theta)) = \frac{\partial}{\partial \mu} \log\left(\sqrt{2\pi\sigma^2} \cdot \frac{1}{2\sigma^2}\right)$$

$$= \frac{\mu}{\sigma^2}$$

$$\frac{\partial}{\partial \mu} \ln(L(\theta)) = \frac{\partial}{\partial \mu} \left(\frac{\mu}{\sigma^2}\right) = \frac{1}{\sigma^2}$$

Applying the thm.

$$\frac{\partial}{\partial \mu} \log\left(\frac{1}{\sigma^2}\right) = E\left(\frac{\partial}{\partial \mu} \ln(L(\theta)) t_i(x)\right) + \left(\frac{\partial}{\partial \mu} \ln(L(\theta))\right) t_i(x)$$

$$\frac{u}{\sigma^2} = \frac{1}{\sigma^2} E(X)$$

$\Rightarrow E(X) = u$.

To obtain $E(\epsilon_2(x)) = E(x^2)$:

$$\begin{aligned} \frac{\partial}{\partial \sigma} \log\left(\frac{1}{\sigma}\right) &= \frac{\partial}{\partial \sigma} \left(\log \sqrt{\frac{1}{\sigma} + \frac{u^2}{\sigma^2}} \right) \\ &= \frac{1}{\sigma} - \left(\frac{u^2}{\sigma^2} \right) \end{aligned}$$

$$\frac{\partial}{\partial \sigma} w_1(\sigma) = \frac{\partial}{\partial \sigma} \left(\frac{1}{\sigma} \right) = \frac{1}{\sigma^2}$$

$$\frac{\partial}{\partial \sigma} w_2(\sigma) = \frac{\partial}{\partial \sigma} \left(\frac{u}{\sigma} \right) = -2u \frac{1}{\sigma^2}$$

Applying them:

$$\frac{\partial}{\partial \sigma} \log\left(\frac{1}{\sigma}\right) = E\left(\frac{1}{\sigma^2} w_1(\sigma) + \left(\frac{\partial}{\partial \sigma} w_2(\sigma)\right) \epsilon_2(x)\right)$$

$$\frac{1}{\sigma} - \frac{u^2}{\sigma^3} = \frac{1}{\sigma^2} E(x^2) - \frac{2u}{\sigma^3} E(X)$$

$$\sigma^2 - u^2 = E(x^2) - 2u^2$$

$$\text{Hence } E(x^2) = \sigma^2 + u^2$$

$$\Rightarrow V(x) = E(x^2) - E(x)^2$$

$$= E(x^2) - u^2$$

$$= \sigma^2$$

3.4: Location and Scale Families

Fact: If $f(x)$ is a valid pdf on \mathbb{R} ,
then $\forall u \in \mathbb{R}, \sigma > 0$,

the function

$$g(x|u, \sigma) := \frac{1}{\sigma} f\left(\frac{x-u}{\sigma}\right), x \in \mathbb{R}$$

is a valid pdf on \mathbb{R} .

Proof: $g \geq 0 \ \forall x \in \mathbb{R}$.

$$\text{Check integral: } \int_{\mathbb{R}} g(x|u, \sigma) dx = \int_{\mathbb{R}} \frac{1}{\sigma} f\left(\frac{x-u}{\sigma}\right) dx$$

Letting $y = \frac{x-u}{\sigma} \Rightarrow x = u + \sigma y$
 $\Rightarrow dx = \sigma dy$

$$= \int_{\mathbb{R}} \frac{1}{\sigma} f(y) \sigma dy = \int_{\mathbb{R}} f(y) dy = 1.$$

Def: Let $f(x)$ be any pdf on \mathbb{R} .

$$\text{The family of pdf } \left\{ g(x|u, \sigma) := \frac{1}{\sigma} f\left(\frac{x-u}{\sigma}\right) \mid u \in \mathbb{R}, \sigma > 0 \right\}$$

is called a location-scale family of distributions for $\begin{cases} u: \text{location param} \\ \sigma: \text{scale param} \end{cases}$

Special cases: 1) $\sum g(x|u) = f(x-u) \quad (u \in \mathbb{R})$ is location family

2) $\left\{ g(x|\sigma) := \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right) \mid \sigma > 0 \right\}$ is scale family

Prop: If $X \sim f$, then $X+u \sim g(x) = f(x-u)$

If $X \sim f$, then $\sigma X \sim g(x) = \frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$

If $X \sim f$, then $\sigma X + u \sim g(x) = \frac{1}{\sigma} f\left(\frac{x-u}{\sigma}\right)$

Ex(1): Let $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} x^2\right)$. Then, the L-S family is:

$$\left\{ g(x|u, \sigma) = \frac{1}{\sigma} f\left(\frac{1}{\sigma}(x-u)\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \left(\frac{x-u}{\sigma}\right)^2\right) \mid u \in \mathbb{R}, \sigma > 0 \right\}$$

Ex(2): Let $f(x) = \frac{1}{\Gamma(a)} x^{a-1} e^{-x}$; $x > 0$

$$\left\{ g(x|u, \sigma) = \frac{1}{\Gamma(a)} \frac{1}{\sigma} \left(\frac{x-u}{\sigma}\right)^{a-1} e^{-\frac{x-u}{\sigma}} \mid u \in \mathbb{R}, \sigma > 0 \right\}$$

Not a Gamma family (incorrect support)

However,

$$\left\{ g(x|u, \sigma) = \frac{1}{\Gamma(a)} \frac{1}{\sigma} \left(\frac{x-u}{\sigma}\right)^{a-1} e^{-\frac{x-u}{\sigma}} \mid u \in \mathbb{R}, \sigma > 0 \right\}$$

$$= \left\{ \text{Gamma}(x, \sigma) \mid u \in \mathbb{R}, \sigma > 0 \right\}$$

Ex(3): Let $f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$, $x=0,1,\dots$ (X ~ Poisson(λ))

Define $g(x|\sigma) = \frac{1}{\sigma} e^{-\lambda} \frac{\lambda^{x+\sigma}}{(x+\sigma)!}$ for $x \geq 0, \sigma \geq 0$

Then, g is a valid pmf supported by multiples of σ .

note: not a Poisson family but inherits many properties of Poisson dist.

Thm: Suppose Y is a RV w/ pmf $f(y)$ and $E(Y)$ exist.

If K is a RV w/ pmf $\frac{1}{\sigma} f\left(\frac{y-\mu}{\sigma}\right)$, then

$$\begin{cases} E(K) = \sigma E(Y) + \mu \\ \text{Var}(K) = \sigma^2 \text{Var}(Y) \end{cases}$$

Proof (sketch): If $Y \sim F$, let $Z := \mu + \sigma Y \sim \frac{1}{\sigma} F\left(\frac{y-\mu}{\sigma}\right)$

and $Z \stackrel{d}{=} K$.

$$\Rightarrow \begin{cases} E(Y) = \mu \\ \text{Var}(Y) = \sigma^2 \end{cases}$$

$$\Rightarrow \begin{cases} E(K) = \mu + \sigma E(Y) \\ \text{Var}(K) = \sigma^2 \text{Var}(Y) \end{cases}$$

3.5: Inequalities and Identities

- when we can't calculate prob's, it is important to estimate them by inequalities to obtain bounds

Thm (Chebyshev's inequality): Let X be a RV, $g(x) \geq 0 \forall x$

$$\text{Then, } P(g(X) \geq r) \leq \frac{E(g(X))}{r} \quad r > 0$$



$$\Leftrightarrow E(g(X)) \geq r \cdot P(g(X) \geq r)$$

$$\text{Proof: } E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_{-\infty}^{\infty} \underbrace{g(x)}_{\geq 0} \underbrace{f_X(x)}_{\geq 0} dx$$

$\geq \int_{r/2}^{\infty} g(x) f_X(x) dx$

$\geq \int_{r/2}^{\infty} r f_X(x) dx$

where $f_X(x) \geq 0$

$$\geq \int_{r/2}^{\infty} r f_X(x) dx$$

$$= r P(g(X) \geq r)$$

and divide both sides by r to conclude. \square

Ex: Let $g(x) = \frac{(x-\mu)^2}{\sigma^2}$ where $\begin{cases} \mu = E(X) \\ \sigma^2 = \text{Var}(X) \end{cases}$

$$\text{Then, } P\left(\frac{(X-\mu)^2}{\sigma^2} \geq 2r\right) \leq \frac{1}{r} E\left(\frac{(X-\mu)^2}{\sigma^2}\right)$$

$$= \frac{1}{r} ?$$

Write $r = t^2$, $t \geq 0$ to obtain

$$\begin{aligned} P((X-\mu)^2 \geq t^2) &\leq \frac{1}{t^2} \\ P((X-\mu)^2 \geq t^2) &\geq \frac{1}{t^2} \end{aligned}$$



$$\text{Thus, } t \geq 2: P((X-\mu)^2 \geq 4) \leq \frac{1}{4} \quad t \geq 1: P((X-\mu)^2 \geq 1) \leq 1$$

$$t \geq 3: P((X-\mu)^2 \geq 9) \leq \frac{1}{9} \quad t \geq 1.1: P((X-\mu)^2 \geq 1.21) \leq 1$$

$$t \geq 4: P((X-\mu)^2 \geq 16) \leq \frac{1}{16} \approx 0.0625$$

...

Thm: (Right tail inequality for normal tails)

If $Z \sim N(0,1)$, then for $t > 0$:

$$P(|Z| \geq t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}$$



Remark: If $t = 2$, $P(|Z| \geq 2) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-2^2/2}}{2} = 0.054$ (Cf. 1/19 from above Chebyshev)

corollary: in general, $\sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t} \leq \frac{1}{t^2}$
exponentially small exponentially large

Proof: $P(|Z| \geq t) = 2P(Z \geq t)$ by symmetry of f_Z

$$\geq 2 \int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$$\leq 2 \int_t^{\infty} \frac{1}{\sqrt{2\pi}} \frac{1}{z^2} e^{-z^2/2} dz$$

$$= \frac{2}{\sqrt{2\pi}} \frac{1}{t} e^{-t^2/2} \Big|_{\infty}^t$$

$$= \sqrt{\frac{2}{\pi}} \frac{1}{t} e^{-t^2/2} \quad \square$$