

Week 10

3-2b: Continuous Distributions

Should find $\sqrt{\pi} = \sqrt{\pi} \int_0^{\infty} 0 \cdot \frac{1}{\sqrt{\pi}} e^{-x^2} dx = \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-x^2} dx$ } Prove/Show

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

Recall: Let $X \sim N(0,1)$. Then $Y = X^2 \sim \chi^2_{df=1}$

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

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

which is pdf for Gamma ($\alpha = \frac{1}{2}, \beta = \frac{1}{2}$)

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

Let $\begin{cases} np \rightarrow \lambda \\ np(1-p) \rightarrow \lambda \end{cases}$. Then $X \xrightarrow{d} \text{Poisson}(\lambda)$.

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

$$\frac{1}{\sqrt{n}}(X - np) \xrightarrow{d} \text{Normal}(0, \sqrt{np(1-p)}) \quad \text{if } \begin{cases} n \rightarrow \infty \\ p \rightarrow p_0 \end{cases}$$

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

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

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

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

$$\text{Then } E(X) = np = 15$$

$$\text{Var}(X) = np(1-p) = 2.45$$

* Poisson approx. not appropriate, why?

now use normal approx:

$$P(X \leq 15) \approx P(Y \leq 15), \quad Y \sim N(15, 2.45)$$

$$= P\left(Z \leq \frac{15-15}{\sqrt{2.45}}\right), \quad Z \sim N(0,1)$$

$$= P(Z \leq 0.00) = 0.506$$

vs. direct computation, i.e.

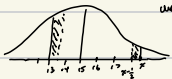
$$P(X \leq 15) = \sum_{k=0}^{15} \binom{15}{k} 0.6^k (0.4)^{15-k} = 0.767$$

Butter yet, use the continuity correction:

$$P(X \leq 15), \quad X \in \mathbb{N}$$

$$\text{and } \approx P(Y \leq 15.5), \quad Y \in \mathbb{R}$$

$$= P\left(Z \leq \frac{15.5-15}{\sqrt{2.45}}\right) = 0.771$$



$$\text{we use } \begin{cases} P(X \leq x) \approx P(Y \leq x + 1/2) \\ P(X \geq x) \approx P(Y \geq x - 1/2) \end{cases}$$

without this correction, we tend to underestimate probs at the tails

3-2c: Continuous Distributions

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

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

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

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

Beta pdfs for different (α, β) :

- $\alpha < \beta = 1 \Rightarrow \text{Beta}(1,1) \equiv \text{Uniform}(0,1)$
- $\alpha > \beta = 1 \Rightarrow \text{Beta}$ is unimodal
- $\alpha < 1, \beta < 1 \Rightarrow \text{Beta}$ is bimodal
- $\alpha < \beta$ or $\alpha > \beta$
- $\alpha > \beta$ or $\alpha < \beta$

(n-1)st moment: $E(X^{n-1}) = \frac{1}{B(a,b)} \int_0^1 x^{n-1} x^{a-1} (1-x)^{b-1} dx$

$$= \frac{1}{B(a,b)} \int_0^1 x^{n+a-1} (1-x)^{b-1} dx$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+n)\Gamma(b)}{\Gamma(a+n+b)}$$

$$= \frac{\Gamma(a+b)\Gamma(a+n)}{\Gamma(a)\Gamma(a+b+n)}$$

$$n! := E(X^n) = \frac{\Gamma(a+b)\Gamma(a+1)}{\Gamma(a)\Gamma(a+b+1)}$$

where $\Gamma(a+1) = a \Gamma(a)$

$$= \frac{\Gamma(a+b) a \Gamma(a)}{\Gamma(a)\Gamma(a+b+1)} = \frac{a}{a+b}$$

Similarly, $E(X^b) = \frac{\Gamma(a+b)\Gamma(b+1)}{\Gamma(a)\Gamma(a+b+1)} = \frac{b}{a+b}$

So, $Var(X) = E(X^2) - E^2(X) = \frac{a(a+b)}{(a+b)^2(a+b+1)} - \left(\frac{a}{a+b}\right)^2$

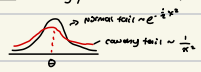
$$= \frac{a}{(a+b)^3} \frac{(a+b)(a+b+1) - a(a+b+1)}{(a+b+1)}$$

$$= \frac{a b}{(a+b)^2 (a+b+1)}$$

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

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

Remarks: Cauchy pdf also bell-shaped:



θ is the median (location) of X

Recall: $\begin{cases} E(X) = \infty \\ E(X) \text{ DNE} \end{cases}$

Prop: Let $X, Y \stackrel{i.i.d.}{\sim} N(0,1)$. Then $\frac{X}{Y} \sim \text{Cauchy}(0)$

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

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

$$f(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} \exp\left(-\frac{1}{2\sigma^2} (\log(x) - \mu)^2\right)$$

3-2: Exponential Families

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

$$f(x|\theta) = c(\theta) h(x) \exp\left\{\sum_{i=1}^k \eta_i(\theta) t_i(x)\right\}, x \in \mathcal{X}$$

$$= c(\theta) h(x) \exp\{u(\theta) \cdot t(x)\} \quad c(\theta)$$

Remarks: where $u(\theta) = (\eta_1(\theta), \dots, \eta_k(\theta))$
 $t(x) = (t_1(x), \dots, t_k(x))$
 \rightarrow vector of sufficient statistics (coming later)

Here,

$h(x) \geq 0$
 $t(x) = (t_1(x), \dots, t_k(x))$ depends only on x (data)
 $u(\theta) = (\eta_1(\theta), \dots, \eta_k(\theta))$ depends only on θ (parameters)

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

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

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

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

$$\theta = p$$

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

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

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

\uparrow \uparrow \uparrow \uparrow
 $c(\theta)$ $h(x)$ $t(x)$ $u(\theta)$ $k=1$

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

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

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

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2)\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2} \exp\left\{\underbrace{\frac{1}{\sigma^2}\mu x}_{\eta(x)} + \underbrace{\frac{\mu^2}{2\sigma^2}}_{\eta_0(x)}\right\}$$

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

Poisson, Geometric, Negative Binomial, Exponential, LogNormal

Among dist.'s not in this (exp.) family are:
Cauchy, Mixtures

Thm: If X is a RV with pdf/pmf in the exponential family of form (6)

then,

$$E(\xi_j) := E\left[\xi_j\right] = \frac{\partial}{\partial \theta_j} \log(1/c(\theta)) = \frac{\partial}{\partial \theta_j} \log(1/c(\theta)), \quad j = 1, \dots, k$$

$$E(\xi_j) := \text{Var}\left(\xi_j\right) = \frac{\partial^2}{\partial \theta_j^2} \log(1/c(\theta)) = \frac{\partial^2}{\partial \theta_j^2} \log(1/c(\theta)) = -E\left[\frac{\partial^2 \log(1/c(\theta))}{\partial \theta_j^2} \xi_j\right] \quad \text{*** useful!$$

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

Corollary (Sufficient): Suppose

$$w(\theta) = \theta = (\theta_1, \theta_2, \dots, \theta_k) \in \mathbb{R}^k$$

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

then $f(x|\theta) = c(\theta) h(x) \exp\langle \theta, t(x) \rangle$ and

$$E(\xi_j) = \frac{\partial}{\partial \theta_j} \log(1/c(\theta))$$

$$E(\xi_j) = \text{Var}(\xi_j) = \frac{\partial^2}{\partial \theta_j^2} \log(1/c(\theta))$$

Proof: proceeds simply via calculus

$$EX(\text{Bernoulli}): f_X(x|\theta) = (1-p)^n \exp\left\{x \log \frac{p}{1-p}\right\}$$

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

$$c(\theta) \quad \eta(x) \quad \eta_0(x) \quad \eta_1(x), \quad k=1$$

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

$$\frac{\partial}{\partial \theta} \log(1/c(\theta)) = -\frac{\partial}{\partial p} \log(1-p) = \frac{n}{1-p}$$

$$\frac{\partial}{\partial \theta} \eta_0(\theta) = \frac{\partial}{\partial p} (1-p) = -1$$

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

Applying the thm:

$$\frac{\partial}{\partial \theta} \log(1/c(\theta)) = E\left[\frac{\partial}{\partial \theta} \log(1/c(\theta)) \xi_j\right]$$

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

$$n = E\left[\frac{x}{p}\right]$$

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

$$EX(\text{Normal}): \text{Given } f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \exp\left\{\underbrace{\frac{1}{\sigma^2}\mu x}_{\eta(x)} + \underbrace{\frac{\mu^2}{2\sigma^2}}_{\eta_0(x)}\right\}$$

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

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

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

$$\frac{\partial}{\partial \mu} \eta_0(\theta) = \frac{\partial}{\partial \mu} \left(\frac{\mu^2}{2\sigma^2}\right) = \frac{\mu}{\sigma^2}$$

Applying the thm:

$$\frac{\partial}{\partial \mu} \log(1/c(\theta)) = E\left[\frac{\partial}{\partial \mu} \log(1/c(\theta)) \xi_1(x) + \left(\frac{\partial}{\partial \mu} \log(1/c(\theta)) \xi_2(x)\right)\right]$$

$$\begin{aligned}\frac{u_1}{\sigma^2} &= \frac{1}{\sigma^2} E(K) \\ &\Rightarrow E(K) = \mu. \\ \text{To obtain } E(t_2(X)) &= E(X^2): \\ \frac{\partial}{\partial \sigma^2} \log\left(\frac{1}{C(\theta)}\right) &= \frac{\partial}{\partial \sigma^2} \left(\log \sqrt{2\sigma^2} + \frac{\mu^2}{2\sigma^2} \right) \\ &= \frac{1}{\sigma} - \mu \frac{1}{\sigma^3} \\ \frac{\partial}{\partial \sigma^2} u_1(\theta) &= \frac{\partial}{\partial \sigma^2} \left(\frac{1}{\sigma^2} \right) = -\frac{1}{\sigma^3} \\ \frac{\partial}{\partial \sigma^2} u_2(\theta) &= \frac{\partial}{\partial \sigma^2} \left(\frac{\mu}{\sigma^2} \right) = -2\mu \frac{1}{\sigma^3}\end{aligned}$$

Applying them:

$$\begin{aligned}\frac{\partial}{\partial \sigma^2} \log\left(\frac{1}{C(\theta)}\right) &= E\left(\frac{\partial}{\partial \sigma^2} u_1(\theta)\right) t_1(X) + \left(\frac{\partial}{\partial \sigma^2} u_2(\theta)\right) t_2(X) \\ \frac{1}{\sigma} - \mu \frac{1}{\sigma^3} &= \frac{1}{\sigma^2} E(X^2) - \frac{2\mu}{\sigma^3} E(K)\end{aligned}$$

$$\begin{aligned}\sigma^2 - \mu^2 &= E(X^2) - 2\mu^2 \\ \text{Hence } E(X^2) &= \sigma^2 + \mu^2 \\ \Rightarrow V(X) &= E(X^2) - E^2(X) \\ &= E(X^2) - \mu^2 \\ &= \sigma^2.\end{aligned}$$

3.4: Location and Scale Families

Fact: If $f(x)$ is a valid pdf on \mathbb{R} ,
 $\mu \in \mathbb{R}$ and $\sigma > 0$,
 the function
 $g(x|\mu, \sigma) := \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right), x \in \mathbb{R}$
 is a valid pdf on \mathbb{R} .

Proof: $g \geq 0$ b/c $f \geq 0$.

$$\begin{aligned}\text{Check integ: } \int_{\mathbb{R}} g(x|\mu, \sigma) dx &= \int_{\mathbb{R}} \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx \\ \text{Letting } y &= \frac{x-\mu}{\sigma} \Rightarrow x = \mu + \sigma y \\ &\Rightarrow dx = \sigma dy \\ &= \int_{\mathbb{R}} \frac{1}{\sigma} f(y) \sigma dy = \int_{\mathbb{R}} f(y) dy = 1.\end{aligned}$$

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

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

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

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

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

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

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

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

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

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

Ex (2): Let $f(x) = \frac{1}{\Gamma(n)} x^{n-1} e^{-x}$, then

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

NB? a Gamma family (incorrect support)

However,

$$\sum g(x|\sigma) = \frac{1}{\Gamma(n)} \frac{1}{\sigma} \left(\frac{x}{\sigma}\right)^{n-1} e^{-x/\sigma} | \sigma \in \mathbb{R}_+, \sum$$

$$\equiv \sum \text{Gamma}(n, \sigma) | \sigma \in \mathbb{R}_+, \sum$$

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

Define $g(x) = \frac{1}{\sigma} e^{-x} \frac{x^{k+1}}{(k+1)!}$ for $x=0,1,\dots$ for $x \in \mathbb{N}$, $\sigma > 0$

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

using it as a Poisson family but inherits many properties of Poisson dist.

Thm. Suppose Y is a RV w/ pdf $f(y)$ and $E(Y), V(Y)$ exist.
If X is a RV w/ pmf $\frac{1}{\sigma} e^{-x} \frac{x^k}{k!}$, then $\begin{cases} E(X) = \sigma E(Y) + \lambda \\ V(X) = \sigma^2 V(Y) \end{cases}$

Proof (sketch): If $Y \sim f$, let $Z := \lambda + \sigma Y \sim \frac{1}{\sigma} e^{-z} \frac{z^k}{k!}$

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

$$\Rightarrow \begin{cases} EX = EZ = \lambda + \sigma EY \\ VZ = VZ = \sigma^2 V(Y) \end{cases}$$

3.5: Inequalities and Identities

When we can't calculate probs., it is important to establish them by inequalities to obtain bounds

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

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

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

$$\text{Proof: } E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_{g(x) \geq r} + \int_{g(x) < r}$$

$$\geq \int_{g(x) \geq r} g(x) f(x) dx$$

where $g(x) \geq r$

$$\geq \int_{g(x) \geq r} r f(x) dx$$

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

and divide both sides by r to conclude. \square



Ex: Let $g(x) = \frac{(x-\mu)^2}{\sigma^2}$ where $\sum_{i=1}^n \mu_i = \mu$ and $\sigma^2 = V(X)$

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

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

With $r = t^2$, $t > 0$ to obtain

$$P(|X-\mu| \geq t\sigma) \leq \frac{1}{t^2}$$

$$P(|X-\mu| \leq t\sigma) \geq 1 - \frac{1}{t^2}$$



$$\text{take } t=2: P(|X-\mu| \geq 2\sigma) \leq 1/4 = 25\%$$

$$t=3: P(|X-\mu| \geq 3\sigma) \leq 1/9 \approx 11\%$$

$$t=4: P(|X-\mu| \geq 4\sigma) \leq 1/16 \approx 6.25\%$$

...

Thm. (tighter 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 $t=2$, $P(|Z| \geq 2) \leq \sqrt{2/\pi} \frac{e^{-2}}{2} \approx 0.054 \approx 5.4\%$ (even tighter Chebyshev)

Corollary: in general, $\sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t} < \frac{1}{t^2}$
exponentially small polynomially small

Proof. $P(|Z| \geq t) = 2P(Z \geq t)$ by symmetry $f_Z = f_{-Z}$

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

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

$$= \frac{2}{\sqrt{2\pi}} \frac{1}{t} e^{-\frac{1}{2}t^2} \left(\frac{1}{t} \right)$$

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