Math 404 Exam 1 is Thurs. Feb. 15. **You are allowed 7 sheets of notes and a calculator.** The exam covers HW1-3, and Q1-3. Numbers refer to types of problems on exam. In this class $\log(t) = \ln(t) = \log_e(t)$ while $\exp(t) = e^t$.

0) Get familiar with the following distributions. For continuous distributions, assume formulas are given on the support, and the support is $x > 0$, unless told otherwise.

a) Exponential(θ)= Gamma($\alpha = 1, \theta$): $f(x) = \dfrac{1}{\theta} e^{-x/\theta}$ where x, $\theta > 0$.

$F(x) = 1 - e^{-x/\theta}$, $E(X) = \theta$, $V(X) = \theta^2$, $E[X \wedge x] = \theta(1 - e^{-x/\theta})$, $e_X(d) = \theta$.

$E(X^k) = \theta^k \Gamma(k+1)$ for $k > -1$. If k is a positive integer, $E(X^k) = \theta^k k!$.

$M(t) = (1-\theta t)^{-1}$, $t < 1/\theta$. $VaR_p(X) = -\theta \ln(1-p)$. $TVaR_p(X) = -\theta \ln(1-p) + \theta$.

b) Gamma(α, θ): $f(x) = \dfrac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\theta}$ where α, θ, and x are positive.

$E(X) = \alpha\theta$, $V(X) = \alpha\theta^2$, $E(X^k) = \dfrac{\theta^k \Gamma(\alpha + k)}{\Gamma(\alpha)}$ for $k > -\alpha$.

$M(t) = (1 - \theta t)^{-\alpha}$ for $t < 1/\theta$.

c) (two parameter) Pareto(α, θ): $f(x) = \dfrac{\alpha \theta^\alpha}{(\theta + x)^{\alpha+1}}$ where α, θ, and x are positive.

$F(x) = 1 - \left(\dfrac{\theta}{x+\theta}\right)^\alpha$, $E(X) = \dfrac{\theta}{\alpha - 1}$ for $\alpha > 1$, V(X) $= \dfrac{\theta^2 \alpha}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$.

$e_X(d) = \dfrac{\theta + d}{\alpha - 1}$, $E(X^k) = \dfrac{\theta^k \Gamma(k+1)\Gamma(\alpha - k)}{\Gamma(\alpha)}$ for $-1 < k < \alpha$.

If $k < \alpha$ is a positive integer, $E(X^k) = \dfrac{\theta^k k!}{(\alpha - 1) \cdots (\alpha - k)}$.

$E[X \wedge x] = \dfrac{\theta}{\alpha - 1}\left[1 - \left(\dfrac{\theta}{x + \theta}\right)^{\alpha - 1}\right]$, for $\alpha \neq 1$, and $E[X \wedge x] = -\theta \ln\left(\dfrac{\theta}{x+\theta}\right)$ for $\alpha = 1$.

$VaR_p(X) = \theta[(1-p)^{-1/\alpha} - 1]$, $TVaR_p(X) = VaR_p(X) + \dfrac{\theta(1-p)^{-1/\alpha}}{\alpha - 1}$ for $\alpha > 1$.

d) If $X \sim$ single parameter Pareto(α, θ): $f(x) = \dfrac{\alpha\theta^\alpha}{x^{\alpha+1}} I(x > \theta)$ where $\alpha > 0$ and θ is real. Note the **support** is $x > \theta$. $F(x) = 1 - \left(\dfrac{\theta}{x}\right)^\alpha$ for $x > \theta$. $E(X) = \dfrac{\alpha\theta}{\alpha - 1}$ for $\alpha > 1$. $V(X) = \dfrac{\alpha\theta^2}{\alpha - 2} - \left(\dfrac{\alpha\theta}{\alpha - 1}\right)^2$ for $\alpha > 2$. $E(X^k) = \dfrac{\alpha\theta^k}{\alpha - k}$ for $k < \alpha$. $E(X \wedge x) = \dfrac{\alpha\theta}{\alpha - 1} - \dfrac{\theta^\alpha}{(\alpha - 1)x^{\alpha - 1}}$ for $x \geq \theta$. $E(X \wedge x) = x$ for $x < \theta$. **Use** $\theta \geq 0$ for loss models.

$VaR_p(X) = \theta[(1-p)^{-1/\alpha}]$, $TVaR_p(X) = \dfrac{\alpha\theta(1-p)^{-1/\alpha}}{\alpha - 1} = VaR_p(X) + \dfrac{1}{\alpha - 1} VaR_p(X)$ for $\alpha > 1$.

e) Uniform(a, b). This distribution has **support** on $a \leq x \leq b$, $f(x) = \dfrac{1}{b-a}$, $F(x) = (x-a)/(b-a)$, $E(X) = (a+b)/2$, $V(X) = (b-a)^2/12$, $e_X(d) = \dfrac{b-d}{2}, 0 \leq a \leq d \leq b$.

f) Weibull(θ, τ): $f(x) = \dfrac{\tau (x/\theta)^\tau e^{-(x/\theta)^\tau}}{x}$ where $\theta > 0$ and $\tau > 0$.

$F(x) = 1 - e^{-(x/\theta)^\tau}$, $E(X^k) = \theta^k \Gamma(1 + k/\tau)$ for $k > -\tau$. Here $\theta, \tau > 0$ and the Weibull$(\theta, \tau = 1)$ RV is the Exponential(θ) RV. $VaR_p(X) = \theta[-\ln(1-p)]^{1/\tau}$.

g) Inverse Weibull(θ, τ): $f(x) = \dfrac{\tau(\theta/x)^\tau e^{-(\theta/x)^\tau}}{x}$.

$F(x) = e^{-(\theta/x)^\tau}$, $E(X^k) = \theta^k \Gamma(1 - k/\tau)$ for $k < \tau$. Here $\theta, \tau > 0$ and the Inverse Weibull$(\theta, \tau = 1)$ RV is the Inverse Exponential(θ) RV. $VaR_p(X) = \theta[-\ln(p)]^{-1/\tau}$.

h) normal(μ, σ): $E(X) = \mu$, $V(X) = \sigma^2$. The **support** is $(-\infty, \infty)$. If $Z \sim N(0,1)$, then the cdf of Z is $\Phi(x)$ and the pdf of Z is $\phi(x)$. If $X \sim N(\mu, \sigma^2)$, then the cdf of X is $F(x) = \Phi\left(\dfrac{x-\mu}{\sigma}\right)$. If $X \sim N(\mu, \sigma^2)$, then the cdf $F_X(x) = \Phi\left(\dfrac{x-\mu}{\sigma}\right)$, and the pdf $f_X(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[\dfrac{-1}{2\sigma^2}(x-\mu)^2\right]$. $TVaR_p(X) = \mu + \sigma \dfrac{\phi(z_p)}{1-p}$ where $P(Z \le z_p) = p$ if $Z \sim N(0,1)$. $VaR_p(X) = \mu + \sigma z_p$. Here $\sigma > 0$ and μ is real.

i) lognormal(μ, σ): $E(X) = \exp(\mu + \frac{1}{2}\sigma^2)$, $V(X) = \exp(\sigma^2)(\exp(\sigma^2) - 1)\exp(2\mu)$, $F(x) = \Phi\left(\dfrac{\ln(x) - \mu}{\sigma}\right)$, $E(X \wedge x) = \exp(\mu + \dfrac{1}{2}\sigma^2)\Phi\left(\dfrac{\ln x - \mu - \sigma^2}{\sigma}\right) + x[1 - \Phi(\dfrac{\ln x - \mu}{\sigma})]$.
If $X \sim LN(\mu, \sigma)$, then $\ln(X) \sim N(\mu, \sigma^2)$. Here $\sigma > 0$ and μ is real.
$VaR_p(X) = \exp(\mu + z_p \sigma)$. For $a > 0$, $aX \sim LN(\mu + \ln(a), \sigma)$.

j) beta(a, b): The **support** is [0,1]. The pdf $f(x) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$ where $a > 0$ and $b > 0$. $E(X) = \dfrac{a}{a+b}$. $V(X) = \dfrac{ab}{(a+b)^2(a+b+1)}$.

The following are discrete distributions. These are used to count the number of claims, so the random variable X is often denoted by N. Note: $p_k = P(X = k) = p(k)$.

k) binomial(q, m): m is a (usually known) positive integer

$$p_k = \binom{m}{k} q^k (1-q)^{m-k} \quad \text{for} \quad \mathrm{k = 0, 1, \ldots, m} \quad \text{where} \quad \mathrm{0 < q < 1}.$$

$E(N) = mq$, $V(N) = mq(1-q)$, $P(z) = [1 + q(z-1)]^m$.

l) Poisson(λ): $p_k = \dfrac{e^{-\lambda}\lambda^k}{k!}$ for $k = 0, 1, \ldots$, where $\lambda > 0$. $E(N) = \lambda = V(N)$, $P(z) = e^{\lambda(z-1)}$.

m) Negative Binomial(β, r): $\beta, r > 0$ and $p_0 = (1+\beta)^{-r}$. For $k = 1, 2, \ldots$,

$$p_k = \frac{r(r+1)\cdots(r+k-1)\beta^k}{k!(1+\beta)^{r+k}} \quad \text{and} \quad \mathrm{p_k = \frac{(k+r-1)!\beta^k}{k!(r-1)!(1+\beta)^{r+k}}} \quad \text{for integer r.}$$

$E(N) = r\beta$, $V(N) = r\beta(1+\beta)$, $P(z) = [1 - \beta(z-1)]^{-r}$. The Geometric$(\beta)$ is the special case with $r = 1$ and $p_k = \dfrac{\beta^k}{(1+\beta)^{k+1}}$ for $k = 0, 1, \ldots$.

Some properties of the gamma function follow.

i) $\Gamma(k) = (k-1)!$ for integer $k \ge 1$.

ii) $\Gamma(x+1) = x\,\Gamma(x)$ for $x > 0$.

iii) $\Gamma(x) = (x-1)\,\Gamma(x-1)$ for $x > 1$.

iv) $\Gamma(0.5) = \sqrt{\pi}$.

Let $X \geq 0$ be a nonnegative random variable.

Then the **cumulative distribution function** (cdf) $F(x) = P(X \leq x)$. Since $X \geq 0$, $F(0) = 0, F(\infty) = 1$, and $F(x)$ is nondecreasing.

The probability density function (**pdf**) $f(x) = F'(x)$.

The **survival function** $S(x) = P(X > x)$. $S(0) = 1, S(\infty) = 0$ and $S(x)$ is nonincreasing.

The **hazard rate function** = *force of mortality* $= \mu(x) = h(x) = \dfrac{f(x)}{1 - F(x)} = \dfrac{f(x)}{S(x)}$ for $x > 0$ and $F(x) < 1$. Note that $h(x) \geq 0$ if $F(x) < 1$.

The **cumulative hazard function** $H(x) = \int_0^x h(t)dt$ for $x > 0$. It is true that $H(0) = 0, H(\infty) = \infty$, and $H(x)$ is nondecreasing.

Assume $X \geq 0$ unless told otherwise.

1) Given one of $F(x), f(x), S(x), h(x)$, or $H(x)$, be able to find the other 4 quantities for $x > 0$. See HW1.

A) $F(x) = \int_0^x f(t)dt = 1 - S(x) = 1 - \exp[-H(x)] = 1 - \exp[-\int_0^x h(t)dt]$.

B) $f(x) = F'(x) = -S'(x) = h(x)[1 - F(x)] = h(x)S(x) = h(x)\exp[-H(x)] = H'(x)\exp[-H(x)]$.

C) $S(x) = 1 - F(x) = 1 - \int_0^x f(t)dt = \int_x^\infty f(t)dt = \exp[-H(x)] = \exp[-\int_0^x h(t)dt]$.

D) $h(x) = \dfrac{f(x)}{1 - F(x)} = \dfrac{f(x)}{S(x)} = \dfrac{F'(x)}{1 - F(x)} = \dfrac{-S'(x)}{S(x)} = -\dfrac{d}{dx}\ln[S(x)] = H'(x)$.

E) $H(x) = \int_0^x h(t)dt = -\ln[S(x)] = -\ln[1 - F(x)]$.

Tip: if $F(x) = 1 - \exp[G(x)]$ for $x > 0$, then $H(x) = -G(x)$ and $S(x) = \exp[G(x)]$.

Tip: For $S(x) > 0$, note that $S(x) = \exp[\ln(S(x))] = \exp[-H(x)]$. Finding $\exp[\ln(S(x))]$ and setting $H(x) = -\ln[S(x)]$ is easier than integrating $h(x)$.

2) **Know:** Except for the inverse Gaussian distribution, the continuous distributions in Appendix A with parameter θ are scale families with scale parameter θ if any other parameters $\boldsymbol{\tau}$ are fixed, written $X \sim SF(\theta|\boldsymbol{\tau})$. Let $a > 0$. Then $Y = aX \sim SF(a\theta|\boldsymbol{\tau})$. If $X \sim LN(\mu, \sigma)$, then $Y = aX \sim LN(\mu + \ln(a), \sigma)$. Often $a = 1 + r$.

3) $X \wedge d = \min(X, d)$ is the limited loss RV. This RV is right censored. The limited expected value $E[X \wedge d] = \displaystyle\int_0^d x f(x)dx + dS(d) = \int_0^d S(x)dx$. The expected loss (per loss) for a policy holder with deductible d is $E[X \wedge d]$.

4) Let $X \geq 0$ be continuous. If $\lim_{x\to\infty} xS(x) = 0$, then $E(X) = \int_0^\infty x f(x)dx = \int_0^\infty S(x)dx = \int_0^\infty [1 - F(x)]dx = \mu =$ mean. The kth raw moment $= \mu'_k = E(X^k) = \displaystyle\int_0^\infty x^k f(x)dx$. If $\lim_{x\to\infty} x^k S(x) = 0$ and $k \geq 1$, then $E(X^k) = \displaystyle\int_0^\infty k x^{k-1} S(x)dx$.

If X is discrete, $= E(X^k) = \displaystyle\sum_k x^k P(X = x)$.

5) The kth central moment $\mu_k = E[(X - \mu)^k]$. The variance uses $k = 2$ and the short cut formula for the variance is $V(X) = E[(X - \mu)^2] = \sigma^2 = E(X^2) - [E(X)]^2$ where $\mu = E(X)$. Note: $\mu_3 = \mu'_3 - 3\mu'_2\mu + 2\mu^3$ and $\mu_4 = \mu'_4 - 4\mu'_3\mu + 6\mu'_2\mu^2 - 3\mu^4$.

The standard deviation $SD(X) = \sqrt{V(X)} = \sigma$.

6) Suppose $X \geq 0$. Then $E[g(X)] = \int_0^\infty g(x) f(x) dx$ for X continuous and $E[g(X)] = \sum_{x:p(x)>0} g(x)\ p(x)$ for X discrete.

7) The coefficient of variation $= CV = \frac{\sigma}{\mu}$, skewness $= \gamma_1 = \frac{\mu_3}{\sigma^3}$, and kurtosis $= \gamma_2 = \frac{\mu_4}{\sigma^4}$. For a statistic T, $CV(T) = SD(T)/E(T)$.

8) The **per loss** RV $Y^L = (X-d)_+ = 0$ if $X \leq d$, $Y^L = (X-d)_+ = X - d$ if $X > d$. The RV is left censored since values $X \leq d$ are not ignored but are set to d. So values of $X - d < 0$ are set to 0. Note that $(X-d)_+$ is the positive part of $X - d$, and represents payment for **insurance with a deductible**. The superscript L represents the "payment," possibly 0, made per loss. $E[(X-d)_+] = e_X(d)[1-F(d)] = e_X(d)S(d) = \int_d^\infty (x-d) f(x) dx = \int_d^\infty S(x) dx = E(Y^L) = E(Y^P) S(d)$.

9) For a given value of $d > 0$ with $P(X > d) > 0$, the excess loss variable or **per payment** RV $Y^P = (X-d)|X > d$. This is a left truncated and shifted RV. The mean excess loss function $e_X(d) = E(Y^P) = E[(X-d)|X > d] = \frac{\int_d^\infty (x-d) f(x) dx}{1 - F(d)} = \frac{\int_d^\infty S(x) dx}{S(d)} = \frac{E(Y^L)}{S(d)}$. The superscript P represents "payment" per payment > 0 actually made (so the loss $> d$).

10) Since insurance with a limit d plus insurance with a deductible d equals full coverage insurance: $X \wedge d \ + \ (X-d)_+ = X$, we get $E[X \wedge d] \ + \ E[(X-d)_+] = E[X]$, and $E[X \wedge d] = E[X] - E[(X-d)_+]$. So $E(Y^L) = E(X) - E(X \wedge d)$.

11) $E[(d-X)_+] = d - E[X \wedge d]$

12) The **Value at Risk** of X at the $100p\%$ level $= 100p$th percentile $VaR_p(X) = \pi_p$ satisfies $F(\pi_p) = P(X \leq \pi_p) = p$ if X is a continuous RV with increasing $F(x)$. Then to find π_p, let $\pi = \pi_p$ and solve $F(\pi) \stackrel{\text{set}}{=} p$ for π.

For a general RV X, π_p satisfies $F(\pi_p-) = P(X < \pi_p) \leq p \leq F(\pi_p) = P(X \leq \pi_p)$. So $F(\pi_p-) \leq p$ and $F(\pi_p) \geq \alpha$. Then graphing $F(x)$ can be useful for finding π_p.

13) The *tail value at risk* of X at $100p\%$ security level is

$$TVaR_p(X) = E(X|X > \pi_p) = \frac{\int_{\pi_p}^\infty x f(x) dx}{1 - F(\pi_p)} = \frac{\int_p^1 \pi_u du}{1-p} = VaR_p(X) + e_X(\pi_p) =$$
$$\pi_p + \frac{\int_{\pi_p}^\infty (x - \pi_p) f(x) dx}{1-p} = \pi_p + \frac{E(X) - E(X \wedge \pi_p)}{1-p}.$$

14) The **loss elimination ratio** LER $= \frac{E[X \wedge d]}{E(X)}$ if $E(X)$ exists. Note that $E(Y^L) = E[(X-d)_+] = E(X) - E[X \wedge d]$. So $E[X \wedge d] = E(X) - E[(X-d)_+] = E(X) - E(Y^L)$.

15) **Know** Given X is a loss RV with parameters $\boldsymbol{\gamma}$, be able to estimate many of the above quantities given $\hat{\boldsymbol{\gamma}}$: plug in $\hat{\boldsymbol{\gamma}}$ for $\boldsymbol{\gamma}$.

16) If there is a policy limit u, then $X \wedge u$ is important. If there is a deductible d, and a maximum payment $u - d$, then $u = u - d + d$.

17) The **method of moments estimator** for a $k \times 1$ parameter vector $\boldsymbol{\gamma}$ sets

$E(X_j) \stackrel{\text{set}}{=} \dfrac{1}{n}\sum_{i=1}^n X_i^j$ for $j = 1, ..., k$ and solves for $\gamma_1, ..., \gamma_k$. The solution is the method of moments estimator $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, ..., \hat{\gamma}_k)$.

In more detail, let $\hat{\mu}_j' = \frac{1}{n}\sum_{i=1}^n X_i^j$, let $\mu_j' = E(X^j)$ and assume that $\mu_j' = \mu_j'(\gamma_1, ..., \gamma_k)$. Solve the system

$$\begin{array}{ccc} \hat{\mu}_1' & \stackrel{\text{set}}{=} & \mu_1'(\gamma_1, ..., \gamma_k) \\ \vdots & & \vdots \\ \hat{\mu}_k' & \stackrel{\text{set}}{=} & \mu_k'(\gamma_1, ..., \gamma_k) \end{array}$$

for the method of moments estimator $\hat{\boldsymbol{\gamma}}$.

18) If g is a continuous function of the first k moments and $h(\boldsymbol{\gamma}) = g(\mu_1'(\boldsymbol{\gamma}), ..., \mu_k'(\boldsymbol{\gamma}))$, then the method of moments estimator of $h(\boldsymbol{\gamma})$ is $g(\hat{\mu}_1', ..., \hat{\mu}_k')$.

19) The method of moments estimator (MME) for $E(X)$ is $\overline{X} = \dfrac{1}{n}\sum_{i=1}^n X_i = m$, the sample mean. The MME for $V(X)$ is the *sample biased variance* = *empirical variance* $= \hat{\sigma}_E^2 = \dfrac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 = \dfrac{1}{n}\sum_{i=1}^n X_i^2 - \left[\dfrac{1}{n}\sum_{i=1}^n X_i\right]^2 = t - m^2$ where $t = \dfrac{1}{n}\sum_{i=1}^n X_i^2$, the sample 2nd moment. Often X_i will be replaced by x_i if $X_1, ..., X_n$ are iid RVs and $x_1, ..., x_n$ are the observed data.

20) The unbiased estimator of the variance is the *sample variance*
$\hat{\sigma}_U^2 = \dfrac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 = \dfrac{n}{n-1}\hat{\sigma}_E^2.$

21) Suppose there are 2 unknown parameters γ_1 and γ_2. Solving $E(X) \stackrel{set}{=} m$ and $E(X^2) \stackrel{set}{=} t$ for γ_1 and γ_2 is equivalent to solving $E(X) \stackrel{set}{=} m$ and $V(X) \stackrel{set}{=} \hat{\sigma}_E^2$ for γ_1 and γ_2: both give the MMEs for γ_1 and γ_2.

22) If there is one unknown parameter γ and $E(X) = g(\gamma)$ where g^{-1} exists (e.g. g is increasing or decreasing), then the MME $\hat{\gamma} = g^{-1}(\overline{X})$.

23) Some useful MMEs where the **parameters are unknown** (except for k in vii)).

i) $G(\alpha, \theta)$: $\hat{\alpha} = \dfrac{m^2}{t - m^2} = \dfrac{m^2}{\hat{\sigma}_E^2}$, $\hat{\theta} = \dfrac{\hat{\sigma}_E^2}{m} = \dfrac{t - m^2}{m}$

ii) EXP(θ): $\hat{\theta} = m$

iii) $U(0, \theta)$: $\hat{\theta} = 2m$

iv) Pareto(α, θ): $\hat{\alpha} = \dfrac{2(t - m^2)}{t - 2m^2} = \dfrac{2\hat{\sigma}_E^2}{t - 2m^2}$, $\hat{\theta} = \dfrac{mt}{t - 2m^2}$

v) LN(μ, σ): $\hat{\mu} = 2\ln(m) - 0.5\ln(t)$, $\hat{\sigma}^2 = \ln(t) - 2\ln(m)$

vi) Poisson(λ): $\hat{\lambda} = m$

vii) binomial(q, k), k **known**: $\hat{q} = \overline{X}/k = m/k$
(the text often uses $k = m$ which can be confusing)

viii) Geometric(β): $\hat{\beta} = m$

ix) negative binomial(r, β): $\hat{r} = \dfrac{m^2}{\hat{\sigma}_E^2 - m} = \dfrac{m}{\hat{\beta}}$, $\hat{\beta} = \dfrac{\hat{\sigma}_E^2 - m}{m} = \dfrac{m}{\hat{r}}$.

24) Suppose X has a mixture distribution where the cdf of X is $F_X(x) = (1 - \epsilon)F_{X_1}(x) + \epsilon F_{X_2}(x)$ where $0 \le \epsilon \le 1$ and F_{X_1} and F_{X_2} are cdfs, then $E[g(X)] =$

$(1-\epsilon)E[g(X_1)] + \epsilon E[g(X_2)]$. In particular, $E(X^2) = (1-\epsilon)E[X_1^2] + \epsilon E[X_2^2] =$ $(1-\epsilon)[V(X_1) + (E[X_1])^2] + \epsilon[V(X_2) + (E[X_2])^2]$. $E(X) = (1-\epsilon)E[X_1] + \epsilon E[X_2]$.

25) If X is a *point mass* at a or *degenerate* at a, then $P(X = a) = 1$. Often $a = 0$.

26) Let $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma)$. *Percentile matching* matches k percentiles instead of k moments. Usually $k = 1$ or 2. Solve the system

$$\begin{array}{ccc} \hat{\pi}_{p_1} & \stackrel{\text{set}}{=} & \pi_{p_1}(\gamma_1, ..., \gamma_k) \\ \vdots & & \vdots \\ \hat{\pi}_{p_k} & \stackrel{\text{set}}{=} & \pi_{p_k}(\gamma_1, ..., \gamma_k) \end{array}$$

for $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, ..., \hat{\gamma}_k)$. Here $F(\pi_{p_j}) = p_j$ and $F(\hat{\pi}_{p_j}) \approx p_j$. $VaR_p(X) = \pi_p$ is given for some brand name distributions. Usually X comes from a continuous distribution.

27) Let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n-1)} \le X_{(n)}$ be the order statistics of $X_1, ..., X_n$. Let the greatest integer function $\lfloor x \rfloor$ = the greatest integer $\le x$, i.e. $\lfloor 7.7 \rfloor = 7$. The *smoothed empirical estimator of a percentile* π_p is $\hat{\pi}_p = X_{(j)}$ if $j = (n+1)p$ is an integer, and $\hat{\pi}_p = (1-h)X_{(j)} + hX_{(j+1)}$ if $(n+1)p$ is not an integer where $j = \lfloor (n+1)p \rfloor$ and $h = (n+1)p - j$. Here $\hat{\pi}_p$ is undefined if $j = 0$ or $j = n+1$, equivalently, $\hat{\pi}_p$ is undefined if $0 \le p < 1/(n+1)$ or if $p = 1$.

28) Given a table of intervals representing loss sizes, number of losses (or proportion of losses), and a distribution X with one parameter γ, be able to estimate γ by matching the $100p_j$th percentile π_{p_j}. Let $n = n_1 + n_2 + \cdots + n_m$ and $p_i = (n_1 + \cdots + n_i)/n$. For a given value of p_j, solve $\hat{\pi}_{p_j} = a_j \stackrel{set}{=} \pi_{p_j} = \pi_{p_j}(\gamma)$ for γ. The solution is $\hat{\gamma}$.

interval	number (or proportion n_i/n)	$\hat{\pi}_{p_i}$
$(a_0, a_1]$	n_1	$a_1 = \hat{\pi}_{p_1}$
$(a_1, a_2]$	n_2	$a_2 = \hat{\pi}_{p_2}$
$(a_2, a_3]$	n_3	$a_3 = \hat{\pi}_{p_3}$
$\vdots$	$\vdots$	$\vdots$
$(a_{m-2}, a_{m-1}]$	n_{m-1}	$a_{m-1} = \hat{\pi}_{p_{m-1}}$
$(a_{m-1}, a_m]$	n_m	

29) The 100pth percentile $VaR_p(X) = \pi_p$ satisfies $F(\pi_p) = P(X \le \pi_p) = p$ if X is a continuous RV with increasing $F(x)$. Then to find π_p, solve $F(\pi_p) \stackrel{\text{set}}{=} p$ for π_p.

For a general RV X, π_p satisfies $F(\pi_p-) = P(X < \pi_p) \le p \le F(\pi_p) = P(X \le \pi_p)$. So $F(\pi_p-) \le p$ and $F(\pi_p) \ge \alpha$. Then graphing $F(x)$ can be useful for finding π_p.

30) **Central Limit Theorem (CLT).** Let $X_1, ..., X_n$ be iid with $E(X) = \mu$ and $V(X) = \sigma^2$. Let the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Hence $\sqrt{n}\left(\dfrac{\overline{X}_n - \mu}{\sigma}\right) = \sqrt{n}\left(\dfrac{\sum_{i=1}^n X_i - n\mu}{n\sigma}\right) = \sqrt{n}\left(\dfrac{S_n - n\mu}{n\sigma}\right) \xrightarrow{D} N(0,1)$.

31) The notation $Y_n \xrightarrow{D} X$ means that for large n we can approximate the cdf of Y_n by the cdf of X. The distribution of $X$is the limiting distribution or asymptotic distribution of Y_n, and the limiting distribution does not depend on n.

32) The notation

$$Y_n \approx N(\theta, \tau^2/n),$$

also written as $Y_n \sim AN(\theta, \tau^2/n)$, means approximate the cdf of Y_n as if $Y_n \sim N(\theta, \tau^2/n)$. Note that the approximate distribution, unlike the limiting distribution, does depend on n. By the CLT, $\overline{X}_n \sim AN(\mu, \sigma^2/n)$ and $S_n = \sum_{i=1}^n X_i \sim AN(n\mu, n\sigma^2)$.

33) Suppose $z_p = \pi_p$ for the N(0,1) distribution: $P(Z \leq z_p) = p$. If $X \sim N(\mu, \sigma^2)$ and π_p is the 100pth percentile of X with $P(X \leq \pi_p) = p$, then $\pi_p = VaR_p(X) = \mu + \sigma z_p$.

If a statistic $T_n \sim AN(\gamma, \psi^2)$. Then use the normal approximation to find i) $P(a < T_n < b) \approx P\left(\frac{a-\gamma}{\psi} < Z < \frac{b-\gamma}{\psi}\right)$ where $<$ can be replaced by $\leq$ unless T_n is discrete and a continuity correction is desired. ii) $\pi_p(T_n) = VaR_P(T_n) \approx \gamma + \psi z_p$. For example, if $T_n = \overline{X}_n$, then $\gamma = \mu$ and $\psi^2 = \sigma^2/n$.

34) Here are some percentile matching formulas if $X_1, ..., X_n$ are iid with distribution X.

a) $X \sim EXP(\theta)$: $\hat{\theta} = \dfrac{-\hat{\pi}_p}{\ln(1-p)}$

b) $X \sim$ Inverse Exponential (θ): $\hat{\theta} = -\hat{\pi}_p \ln(p)$

c) $X \sim LN(\mu, \sigma)$: $\hat{\mu} = \ln(\hat{\pi}_p) - z_p\hat{\sigma}$, $\hat{\sigma} = \dfrac{\ln(\hat{\pi}_p) - \ln(\hat{\pi}_q)}{z_p - z_q}$

d) $X \sim$ Weibull(θ, τ): $\hat{\theta} = \dfrac{\hat{\pi}_p}{[-\ln(1-p)]^{1/\hat{\tau}}}$, $\hat{\tau} = \dfrac{\ln[\ln(1-p)/\ln(1-q)]}{\ln(\hat{\pi}_p/\hat{\pi}_q)}$

35) For right censored data $X_1, ..., X_m$, $n - m$ cases censored at $u > X_{(m)}$, the order statistics are $X_{(1)}, ..., X_{(m)}, u, ..., u$. If $j + 1 \leq m$, then percentile matching can still be used with $\hat{\pi}_p$ from 27).

36) If X is (left) truncated at d then $W = X|X > d$ has survival function $S_W(x) = \dfrac{S_X(x)}{S_X(d)}$ for $x > d$, and cdf $F_W(x) = 1 - S_W(x)$ for $x > d$. If data is iid from the truncated distribution, e.g. if the losses include the deductible d, find $\hat{\pi}_p$ as in 27), but solve $\dfrac{S_X(\hat{\pi}_p)}{S_X(d)} \stackrel{set}{=} 1 - p$ for γ. Use two equations with $\hat{\pi}_p$ and $\hat{\pi}_q$ if you need to estimate two parameters γ_1 and γ_2. (The brand name distribution X is being fit, but you have left truncated data at d, so the equations for percentile matching are changed.)

37) Let $h(x) \equiv h_X(x|\boldsymbol{\theta})$ be the pdf or pmf of a random variable X. Let the set Θ be the set of parameter values $\boldsymbol{\theta}$ of interest. Then the set $\mathcal{X}_{\boldsymbol{\theta}} = \{x | h_Y(x|\boldsymbol{\theta}) > 0\}$ is called the *sample space* or **support** of X, and Θ is the **parameter space** of X. Often $\Theta = \{\boldsymbol{\theta} | h(x|\boldsymbol{\theta})$ is a pdf or pmf $\}$. Use the notation $\mathcal{X} = \{x | h(x|\boldsymbol{\theta}) > 0\}$ if the support does not depend on $\boldsymbol{\theta}$. So $\mathcal{X}$ is the support of X if $\mathcal{X}_{\boldsymbol{\theta}} \equiv \mathcal{X} \ \forall \boldsymbol{\theta} \in \Theta$. Similar definitions can be used for $\boldsymbol{X} = (X_1, ..., X_n)$.

38) Let $\boldsymbol{X} = (X_1, ..., X_n)$. If $\boldsymbol{x} = (x_1, ..., x_n)$ is the data then the **likelihood function** $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}|\boldsymbol{x})$. For each sample point $\boldsymbol{x} = (x_1, ..., x_n)$, let $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ be a parameter value at which $L(\boldsymbol{\theta}|\boldsymbol{x})$ attains its maximum as a function of $\boldsymbol{\theta}$ with $\boldsymbol{x}$ held fixed. Then a maximum likelihood estimator (**MLE**) of the parameter $\boldsymbol{\theta}$ based on the sample $\boldsymbol{X}$ is $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$. Note: it is crucial to observe that the likelihood function is a function of $\boldsymbol{\theta}$ (and that $x_1, ..., x_n$ act as fixed constants).

39) If the MLE $\hat{\boldsymbol{\theta}}$ exists, then $\hat{\boldsymbol{\theta}} \in \Theta$. If the MLE $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, ..., \hat{\theta}_k)$, then the MLE of θ_i is $\hat{\theta}_i$, the MLE of (θ_1, θ_5) is $(\hat{\theta}_1, \hat{\theta}_5)$, etc.

40) **Invariance Principle:** If $\hat{\boldsymbol{\theta}}$ is the MLE of θ, then $\tau(\hat{\boldsymbol{\theta}})$ is the MLE of $\tau(\boldsymbol{\theta})$. Here τ is a function of $\boldsymbol{\theta}$ with domain Θ.

41) For **individual data**, $X_1, ..., X_n$ are iid, usually with pdf $f(x)$ or pmf $p(x)$. Let $\boldsymbol{x} = (x_1, ..., x_n)$ be the observed data. Then the **likelihood function** $L(\boldsymbol{\theta}) \equiv L(\boldsymbol{\theta}|\boldsymbol{x}) = \prod_{i=1}^n h(x_i)$ where $h(x)$ is $f(x)$ or $p(x)$. The **log likelihood function** $\ln(L(\boldsymbol{\theta})) = \sum_{i=1}^n \ln(h(x_i))$. Usually use 42) to find the MLE.

42) For this class, assume that the maximum likelihood estimator (MLE) is a solution to $\dfrac{\partial}{\partial \theta_i} \ln L(\boldsymbol{\theta}) \stackrel{set}{=} 0$ for $i = 1, ..., k$ where usually $k = 1$ or 2. (In Math 483 or 580, used second derivatives to show that the MLE was the global max.)

Tips: a) $\exp(a) = e^a$. b) $\ln(a^b) = b\ln(a)$ and $\ln(e^b) = b$. c) $\ln(\prod_{i=1}^n a_i) = \sum_{i=1}^n \ln(a_i)$. d) Often $\ln[L(\theta)] = \ln(\prod_{i=1}^n f(x_i|\theta)) = \sum_{i=1}^n \ln(f(x_i|\theta))$. e) If t is a differentiable function and $t(\theta) \neq 0$, then $\frac{d}{d\theta}\ln(|t(\theta)|) = \frac{t'(\theta)}{t(\theta)}$ where $t'(\theta) = \frac{d}{d\theta}t(\theta)$. In particular, $\frac{d}{d\theta}\ln(\theta) = 1/\theta$. f) Anything that does not depend on θ is treated as a constant with respect to θ and hence has derivative 0 with respect to θ.

43) For small n, if given $\boldsymbol{x}$ it can be easier to plug in the x_i to find the MLE. Sometimes you will solve for the MLE as a statistic, then plug $\boldsymbol{x}$ into the statistic.

44) Let $h(\boldsymbol{x}|\boldsymbol{\theta})$ be the pmf or pdf of a sample $\boldsymbol{X}$. If $\boldsymbol{X} = \boldsymbol{x}$ is observed, then **the likelihood function** $L(\boldsymbol{\theta}) = h(\boldsymbol{x}|\boldsymbol{\theta})$.

45) Let $X_1, ..., X_n$ be iid with distribution X. Here are some MLEs.

a) If $X \sim N(\mu, \sigma^2)$, then the MLE of μ is $\overline{X}$. If μ and σ^2 are unknown, then the MLE of σ^2 is the empirical variance (= method of moments estimator of $V(X)$) $\hat{\sigma}^2 = \hat{\sigma}^2_E = \dfrac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$. If μ is **known**, the MLE of σ^2 is $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$.

b) If $X \sim$ Poisson(λ) then $\hat{\lambda} = \overline{X}$.

c) If $X \sim$ binomial(q, k), k **known**, then $\hat{q} = \overline{X}/k = m/k$.

d) If $X \sim EXP(\theta)$, then $\hat{\theta} = \overline{X}$.

e) If $X \sim$ negative binomial (r, β), the MLE of $r\beta = \overline{X}$, but the MLEs of r and β need a computer. If r is **known**, then $\hat{\beta} = \dfrac{\overline{X}}{r}$.

f) If $X \sim G(\alpha, \theta)$ with α **known**, the MLE of θ is $\overline{X}/\alpha$.

g) If $X \sim$ geometric(β), the MLE of β is $\overline{X}$.

h) If $X \sim LN(\mu, \sigma)$, let $W_i = \ln(X_i)$. Then the MLE of μ is $\overline{W}$. If μ and σ^2 are unknown, then the MLE of σ^2 is the empirical variance of the W_i: $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^n (W_i - \overline{W})^2$.

If μ is **known**, the MLE of σ^2 is $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^n (W_i - \mu)^2$.

i) If $X \sim U(0, \theta)$, the MLE of θ is $\hat{\theta} = X_{(n)}$.

j) If $X \sim$ inverse exponential (θ), then the MLE $\hat{\theta} = \dfrac{n}{\sum_{i=1}^n \frac{1}{X_i}}$.

46) Note that for the $G(\alpha, \beta)$ with α known, binomial(q, k) with k known, EXP(θ), geometric(β), and Poisson(λ) distributions, the MLEs are the same as the MMEs.