Math 404 Exam 3 is Wed. April 27. **You are allowed 20 sheets of notes and a calculator.** The exam covers HW7-10, and Q7-10. The final is Monday, May 9, 12:30-2:30, and is cumulative.**You are allowed 25 sheets of notes and a calculator for the final.**

96) Let $F_n(x)$ be the empirical cdf and let $F^*(x)$ be the fitted cdf. Let $X_1, ..., X_n$ be iid or possibly truncated or censored. Let d be the truncation point ($d = 0$ for no truncation) and let u be the censoring point ($u = \infty$ for no censoring). The Kolmogorov Smirnov test statistic $D = \max_{d \le x \le u} |F_n(x) - F^*(x)| = \max_{x_i}(|F_n(x_i) - F^*(x_i)|, |F_n(x_i-) - F^*(x_i)|)$ where $F(x-) = P(X < x)$. Note that $F_n(x) = \frac{\# x_i \le x}{n}$ and $F_n(x-) = \frac{\# x_i < x}{n}$. Let $x_1 \le x_2 \le ... \le x_n$ be the observed order statistics. If there are no ties, then $F_n(x_i) = i/n$ and $F_n(x_i-) = (i-1)/n$. The following table works when there are no ties. Then D is the largest value in the last column.

x_i	$F_n(x_i)$	$F_n(x_i-)$	$F^*(x_i)$	$\max(\lvert F_n(x_i) - F^*(x_i)\rvert, \lvert F_n(x_i-) - F^*(x_i)\rvert)$
x_1	$1/n$	$0/n$	$F^*(x_1)$	$\max(\lvert 1/n - F^*(x_1)\rvert, \lvert 0/n - F^*(x_1)\rvert)$
x_2	$2/n$	$1/n$	$F^*(x_2)$	$\max(\lvert 2/n - F^*(x_2)\rvert, \lvert 1/n - F^*(x_2)\rvert)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_j	j/n	$(j-1)/n$	$F^*(x_j)$	$\max(\lvert j/n - F^*(x_j)\rvert, \lvert (j-1)/n - F^*(x_j)\rvert)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	n/n	$(n-1)/n$	$F^*(x_n)$	$\max(\lvert n/n - F^*(x_n)\rvert, \lvert (n-1)/n - F^*(x_2)\rvert)$

97) Kolmogorov Smirnov critical values

α	0.1	0.05	0.01
	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

98) 4 step Kolmogorov Smirnov test
i) H_0: fitted distribution is good H_1: not H_0
ii) D
iii) reject H_0 if $D >$ critical value, otherwise fail to reject H_0
iv) non technical conclusion: reject H_0: fitted distribution is not good, fail to reject H_0: fitted distribution is good (or there is not enough evidence to conclude that the fitted distribution is not good).

99) The Anderson Darling test is a competitor of the χ^2 test and the Kolmogorov Smirnov test.

```
     Kolmogorov Smirnov        Anderson Darling      chisquare test
i)   indiv data                indiv data              indiv or grouped data
ii)  contin fits               contin fits          contin or discrete fits
iii) lower the critical  lower the critical         no adjustment for critical
     value if u < oo      value if u < oo           value if u < oo
iv) lower the critical  lower the critical         df = k - r -1 adjusts for
     value if r > 0      value if r > 0             r > 0
```

```
v)crit value decreases   crit value decreases     critical value free of the
   as n increases        as n increases           sample size n
vi) no discretization   no discretization      discretization with group data
vii) uniform weight on  higher weight on the   higher weight on intervals with
all parts of the distr  tails of the distr    lower prob (often the right tail)
```

100) Likelihood Ratio Test (LRT) where H_0: distribution is from model A and H_1: distribution is from model B where **model A is a special case of model B**: Let Θ_0 be the parameter space for H_0 (model A) and let Θ_1 be the parameter space for H_1 (model B). Let $\hat{\boldsymbol{\theta}}_0$ be the MLE for model A where $\hat{\boldsymbol{\theta}}_0 \in \Theta_0$, and let $\hat{\boldsymbol{\theta}}_1$ be the MLE for model B where $\hat{\boldsymbol{\theta}}_1 \in \Theta_1$. Let $L_0 = L(\hat{\boldsymbol{\theta}}_0)$ and $L_1 = L(\hat{\boldsymbol{\theta}}_1)$. then the LRT test statistic is

$$T = -2\ln(\frac{L_0}{L_1}) = 2\ln(\frac{L_1}{L_0}) = 2[\ln(L_1) - \ln(L_0)].$$

Let $df = d = d_B - d_A =$ number of free parameters in $B -$ number of free parameters in A, where a free parameter is not specified, so must be estimated using the MLE. Reject H_0 if $T > \chi^2_{d,1-\alpha} = \chi^2_{d,P}$ on the χ^2 table, otherwise fail to reject H_0. Then the 4 step test is

i) H_0 data is from distribution A H_1: data is from distribution B
ii) T
iii) reject H_0 if $T > \chi^2_{d,P}$, otherwise fail to reject H_0
iv) nontechnical conclusion: reject H_0: data came from distribution B, fail to reject H_0: data came from distribution A (or there is not enough evidence to conclude that the data came from distribution B)

101) LRT if model A is not a special case of model B or if there are models $A_1, A_2, ..., A_k$: Select, for every number of parameters, the model with the highest loglikelihood. Suppose $\alpha = 0.05$ is the significance level. In order to prefer the best 2 parameter model over the best 1 parameter model, need $2(\ln L_2 - \ln L_1) \geq \chi^2_{1,0.95} = 3.841$. If the best 2 parameter model is not good, need $2(\ln L_3 - \ln L_1) \geq \chi^2_{2,0.95} = 5.991$, and so on. If the 2 parameter model is preferred, then start over comparing the 3, 4, ... parameter models with the 2 parameter model. So need $2(\ln L_3 - \ln L_2) \geq \chi^2_{1,0.95} = 3.841$, and if the 3 parameter model is not good, need $2(\ln L_4 - \ln L_2) \geq \chi^2_{2,0.95} = 5.991$, and so on. If $-$loglikelihood is given, multiply the values by -1. See first 2 columns of the table below. The third column is usually omitted.

number of parameters	maximal loglikelihood	Schwarz Bayesian
1	$\ln(L_1)$	$\ln(L_1) - \frac{1}{2}\ln(n)$
2	$\ln(L_2)$	$\ln(L_2) - \frac{2}{2}\ln(n)$
$\vdots$	$\vdots$	
k	$\ln(L_k)$	$\ln(L_k) - \frac{k}{2}\ln(n)$

102) For the above table, the Schwarz Bayesian criterion says take the model that maximizes $\ln(L_r) - \frac{r}{2}\ln(n)$ where n is the sample size and r is the number of parameters. So take the model that maximizes the 3rd column.

103) For Bayesian statistics θ is a random variable. Let $\pi(\theta)$ be the prior pdf or pmf. Let $f(\boldsymbol{x}|\theta)$ be the conditional pdf or pmf: the likelihood function where usually $f(\boldsymbol{x}|\theta) = \prod_{i=1}^n f(x_i|\theta)$. The joint pdf or pmf is $f(\boldsymbol{x}, \theta) = \pi(\theta) f(\boldsymbol{x}|\theta)$. The posterior pdf or pmf is

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}, \theta)}{f(\boldsymbol{x})}.$$

Then the marginal or unconditional pdf or pmf is $f(\boldsymbol{x}) = \int f(\boldsymbol{x}, \theta) d\theta$ if θ has interval support or $f(\boldsymbol{x}) = \sum_\theta f(\boldsymbol{x}, \theta)$ if θ has a pmf.

104) Typically if the prior is a pdf then so is the posterior, and if the prior is a pmf then so is the posterior. The prior will be a pdf if θ is modeling an interval (e.g., like a probability on [0, 1]), and the prior will be a pmf if θ is modeling a countable number of values (e.g., only probabilities 0.3 and 0.7 are of interest).

105) Suppose $\theta = (\theta_1, ..., \theta_k$ is a random vector. **Bayes' Theorem**: for a posterior pdf,

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{x}|\boldsymbol{t})\pi(\boldsymbol{t}) d\boldsymbol{t}}$$

while for a posterior pmf,

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\sum_{\boldsymbol{t}} f(\boldsymbol{x}|\boldsymbol{t})\pi(\boldsymbol{t})}.$$

In both denominators, $\boldsymbol{t}$ is often replaced by $\boldsymbol{\theta}$.

106) **Know:** the posterior pdf or pmf $\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto \pi(\boldsymbol{\theta}) f(\boldsymbol{x}|\boldsymbol{\theta})$, the product of the prior and the likelihood. All constants that do not depend on $\boldsymbol{\theta}$ can be discarded on the right hand side. Then recognize that the right hand side is a brand name distribution or use the fact that a pdf integrates to 1 and a pmf sums to 1: integrate to get the constant c that makes the posterior a pdf (or sum to get c for a pmf).

107) The posterior distribution is a perfectly good probability distribution. Let $W = \theta|\boldsymbol{x}$. Then $P(a < W < b) = P(a < \theta < b|\boldsymbol{x})$ and $E(W) = E(\theta|\boldsymbol{x})$.

108) The posterior support is a subset of the prior support. So if the prior support is (a,b), then the posterior support is a subset of (a,b) (often equal to (a,b)). If the prior support is $\{0.3, 0.5, 0.7\}$, then the posterior support is a subset (often proper) of $\{0.3, 0.5, 0.7\}$.

109) If a conjugate prior is used, then the posterior distribution has the same distribution as the prior distribution, but with different parameters.

110) a) $\theta|\boldsymbol{x} \sim \text{beta}(a, b)$ if $\pi(\theta|\boldsymbol{x}) \propto \theta^{a-1}(1-\theta)^{b-1}$ where $a, b > -1$ and $\theta \in [0, 1]$.

b) $\theta|\boldsymbol{x} \sim \text{gamma}(\alpha, \beta)$ if $\pi(\theta|\boldsymbol{x}) \propto \theta^{\alpha-1} e^{-\theta/\beta}$ where $\alpha, \beta, \theta > 0$.

c) $\theta|\boldsymbol{x} \sim$ single parameter Pareto (α, β) if $\pi(\theta|\boldsymbol{x}) \propto \theta^{-(\alpha+1)}$ where $\theta > \beta$, $\alpha > 0$ and β is real.

d) $\theta|\boldsymbol{x} \sim$ Pareto (α, β) if $\pi(\theta|\boldsymbol{x}) \propto (\beta + \theta)^{-(\alpha+1)}$ where $\alpha, \beta, \theta > 0$.

e) $\theta|\boldsymbol{x} \sim N(\mu, \sigma^2)$ if $\pi(\theta|\boldsymbol{x}) \propto \exp(\left(\frac{-1}{2\sigma^2}(\theta - \mu)^2\right)$ where $\sigma^2 > 0$ and θ and μ are real.

Note that θ takes the place of x and β often takes the place of θ compared to the distributions given on p. 1-2 of the exam 1 review.

111) Let X_{n+1} be a future value of the data given $X_1 = x_1, ..., X_n = x_n$ have been collected. The predictive density (pdf or pmf) $f(x|\boldsymbol{x}) = f(y|\boldsymbol{x}) = f(x_{n+1}|\boldsymbol{x}) =$

$\int f(y|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta} = \int f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta}$ is the updated unconditional (marginal) pdf (or pmf) for X_{n+1} given the data $\boldsymbol{x}$. here $f(x|\boldsymbol{\theta})$ is the likelihood if $n = 1$ and usually $\boldsymbol{\theta} = \theta$.

112) Using some bad notation, $E(X_{n+1}) = E(X_{n+1}|\boldsymbol{x}) = E[E(X_{n+1}|\Theta)|\boldsymbol{x}]$ is the Bayesian premium $= \int x f(x|\boldsymbol{x})dx$ using the predictive density from 111).

113) The Bayesian estimator or Bayes estimator minimizes the expected posterior loss function.

a) For the (mean) square error loss function, $l(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$, the Bayesian point estimator is the mean of the posterior distribution: $\hat{\theta} = E(\Theta|\boldsymbol{x})$.

b) For the absolute value of the error loss function, $l(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$, the Bayesian point estimator is the median of the posterior distribution: $\hat{\theta} = \pi_{0.5}$.

c) For the zero–one loss function, ($l(\hat{\theta}, \theta) = 0$ if $\hat{\theta} = \theta$, and $l(\hat{\theta}, \theta) = 1$ or any constant k if $\hat{\theta} \neq \theta$,) the Bayesian point estimator $\hat{\theta}$ is the mode of the posterior distribution.

114) The sample space S is partitioned into n subsets $A_1, A_2, ..., A_n$ if a) $A_i \cap A_j = \emptyset$ for $i \neq j$, b) $A_i \neq \emptyset$ for $i = 1, ..., n$, and c) $A_1 \cup A_2 \cup \cdots \cup A_n = S$. Often $n = 2$, and A and the complement $\overline{A}$ form a partition of S. Let $A_1, A_2, ..., A_n$ partition S, and let E be an event in S, then

a) $P(E) = P(A_1)P(E|A_1) + P(A_2)P(E|A_2) + \cdots + P(A_n)P(E|A_n)$ and

b) **Bayes' rule:** $P(A_j|E) = \dfrac{P(A_j \cap E)}{P(E)} = \dfrac{P(A_j)P(E|A_j)}{P(E)}$

$$= \frac{P(A_j)P(E|A_j)}{P(A_1)P(E|A_1) + P(A_2)P(E|A_2) + \cdots + P(A_n)P(E|A_n)}.$$

In particular, if $n = 2$, $P(E) = P(A)P(E|A) + P(\overline{A})P(E|\overline{A})$ and

$P(A|E) = \dfrac{P(A)P(E|A)}{P(A)P(E|A) + P(\overline{A})P(E|\overline{A})}$.

In a **Bayes' rule** story problem, 2 or more unconditional probabilities are given (or easy to find with the complement rule). Several conditional probabilities are also given (or easy to find with the complement rule). **Make a tree diagram** with the events corresponding to the unconditional events labelling the left branches and the events corresponding to the conditional probabilities labelling the right branches. Above the left branches place the unconditional probabilities and above the right branches place the conditional probabilities. You will be asked to find an unconditional right branch probability and to use Bayes' rule to find P(left branch | right branch).

Tips: the hard conditional probability, P(left branch | right branch), usually appears at the end of the story problem. This tells you how to label the left branches and the right branches of the tree. (The easy conditional probabilities, P(right branch | left branch), can also tell you how to label the tree.) The probabilities of the left branch sum to one. Each subtree of right branches has probabilities that sum to one. Occasionally you are asked to find both a P(right branch | left branch) (directly from the tree) and P(left branch | right branch) (using Bayes rule).

115) Let $W = X_{n+1} (= X_{n+1}|\boldsymbol{x})$. Often want the Bayesian estimate = posterior mean of $g(\theta_i) = h(W|\theta_i)$. For example, the expected value of the next claim = E(claim $|\boldsymbol{x}$) has $g(\theta_i) =$ E(claim $|\theta_i$) and $P(X_{n+1} > c|\boldsymbol{x})$ has $g(\theta_i) = P(X_{n+1} > c|\theta_i)$. Using the predictive distribution from 111), the Bayesian estimate is $E[g(\Theta)|\boldsymbol{x})] = \sum_i g(\theta_i)\pi(\theta_i|\boldsymbol{x}) = \sum_i h(w|\theta_i)\pi(\theta_i|\boldsymbol{x})$. Replace the sum by an integral if the posterior is a pdf instead of a pmf.

116) Bayesian credibility puts a prior on classes of risks. Let θ_i correspond to class i. Then X = losses follow a different distribution for each class. Note that X is a generic RV: might want the aggregate loss $X = S = \sum_{i=1}^N X_i$. Use the following table to find the Bayesian premium. Enough information needs to be given to find row i) and ii). Often $k = 2$, and if there are j times as many people in class 1 as in class 2, then π(class 1) $= j/(j+1)$ while π(class 2) $= 1/(j+1)$. Each middle row iii) term is a product of the corresponding terms from rows i) and ii). For row iv), the posterior is the ratio of a row iii) term and the row iii) sum. For row v), the hypothetical mean is the conditional mean of each class. The sum of the row vi) terms is the Bayesian premium = predicted expected value, the quantity that you want to find.

row	class 1	...	class k	sum
i) prior	π(class 1)	...	π(class k)	1
ii) likelihood	$f(\boldsymbol{x}$|class 1)	...	$f(\boldsymbol{x}$|class k)	
iii) joint prob	π(class 1)$f(\boldsymbol{x}$|class 1)	...	π(class k)$f(\boldsymbol{x}$|class k)	denom. of Bayes' th.
iv) posterior	$\dfrac{\pi(class\ 1)f(\boldsymbol{x}\vert class\ 1)}{\text{row iii) sum}}$	...	$\dfrac{\pi(class\ k)f(\boldsymbol{x}\vert class\ k)}{\text{row iii) sum}}$	
	$= \pi$(class 1|$\boldsymbol{x}$)	...	$= \pi$(class k|$\boldsymbol{x}$)	1
v)hyp. mean	$\mu_1 = E(X$|class 1)	...	$\mu_k = E(X$|class k)	
vi) B. prem. contr.	$\mu_1\pi$(class 1|$\boldsymbol{x}$)	...	$\mu_k\pi$(class k|$\boldsymbol{x}$)	Bayesian premium

117) Let $S = \sum_{i=1}^N X_i$ where $S = 0$ if $N = 0$. The distribution of S is called a compound distribution with N the primary distribution and X the secondary distribution. Assume the X_i are iid and $X_i \perp\!\!\!\perp N$ unless told otherwise: then $E(S) = E(N)E(X)$ and $V(S) = E(N)V(X) + [E(X)]^2V(N)$.

118) Classical credibility = limited fluctuation credibility. Let M be the underlying manual rate or pure premium. Let X_j be the claims, or losses, or aggregate losses in past experience period j. Let the policyholder experience $(X_1, ..., X_n)$ be the data where n is the number of time periods exposed to a risk. Assume the X_j are independent with $E(X_j) = E(X)$ and $V(X_j) = V(X)$. Let the **coefficient of variation** $CV(X_j) = CV(X) = \dfrac{SD(X)}{E(X)} = \dfrac{\sqrt{V(X)}}{E(X)}$. Let the **credibility premium** $P_C = Z\overline{X} + (1-Z)M$ where the **credibility (factor)** $Z \in [0,1]$. **Full credibility** occurs if $Z = 1$ so $P_C = \overline{X}$. Partial credibility occurs when $Z < 1$. Want to establish credibility standards based on 2 parameters: a) the probability of being in an interval like a CI: 0.9 , 0.95 or 0.99, with z_p given by 1.645, 1.96, or 2.576, and b) the maximum amount of fluctuation to allow: eg $k = 0.05$.

119) Let e be the amount of exposure needed for full credibility (for $Z = 1$). For a

general RV W, want "CI" $eE(W) \pm z_p\sqrt{eV(W)}$ and want the fluctuation $\frac{z_p\sqrt{eV(W)}}{eE(W)} \leq k$ so $1 \pm$ fluctuation $\in [1-k, 1+k]$. Then $e = \left(\frac{z_p}{k}\right)^2 \left(\frac{SD(W)}{E(W)}\right)^2 = \left(\frac{z_p}{k}\right)^2 [CV(W)]^2$ is the **general formula for full credibility**.

120) In the table below, i) for exposure units $e = e_F$: the measurement unit is the (expected) number of exposures where an exposure unit is a) a risk over a time period (eg number of person years) for both number of claims and aggregate losses, b) a claim for severity (claim size). So e_F is the (expected) number of risks needed for full credibility. ii) For number of claims $e = n_F$ is the (expected) number of claims needed for full credibility. Then $e_F = n_F/\lambda$, the (expected) number of claims divided by the expected number of claims per risk λ. iii) For aggregate losses, the exposure unit $e = a_F = n_F E(X)$, the (expected) number of claims times the expected losses per claim. Note that $e_F = a_F/(\lambda E(X))$. So if $W = N$ or $W = S$, then $e = e_F$ while if $W = X$, then $e = n_F$. Want W to be within 100 k% of the expected 100 p% of the time.

121) Suppose $N \sim Pois(\lambda)$. The general formula for e for full credibility is given below for various W. Let $n_0 = \left(\frac{z_p}{k}\right)^2$. Want how many exposures e are needed for full credibility.

experience expressed in	Number of claims $W = N$	claim size (severity) $W = X$	Aggregate losses (pure premium) $W = S$
exposure units $e = e_F$	$\frac{n_0}{\lambda}$	$\frac{n_0}{\lambda}[CV(X)]^2$	$\frac{n_0}{\lambda}(1 + [CV(X)]^2)$
number of claims $e = n_F$	n_0	$n_0[CV(X)]^2$	$n_0(1 + [CV(X)]^2)$
aggregate losses $e = a_F$	$n_0 E(X)$	$n_0 E(X)[CV(X)]^2$	$n_0 E(X)(1 + [CV(X)]^2)$

122) $P_C = M + Z(\overline{X} - M)$. The credibility factor for i) $e < e_F$ exposure units is $Z = \sqrt{e/e_F}$. The credibility factor for $n < n_F$ expected claims is $Z = \sqrt{n/n_F}$, and the credibility factor for $a < a_F$ aggregate claims is $Z = \sqrt{a/a_F}$. If the prior estimate or manual rate or pure premium M is given for what a statistic T estimates, (eg $T = \overline{X}$: the average claim, or $T = \sum X_i$: the total loss), then $P_C = M + Z(T - M)$.

123) Assume all relevant expectations exist. Then iterated expectations or the conditional mean formula is $E(X) = E[E(X|Y)] = E_Y[E_{X|Y}(X|Y)]$. The conditional variance formula is $V(X) = E[V(X|Y)] + V[E(X|Y)]$. Also, $E(X^k) = E[E(X^k|Y)]$ and $E(h(X,Y)) = E(E[h(X,Y)|Y])$.

124) Let $\Theta = \theta_i$ correspond to risk class i. Let $E(X) = E[E(X|\Theta)]$, $EPV = E[V(X|\Theta)]$ and $VHM = V[E(X|\Theta)]$. Note that $V(X) = EPV + VHM$ by 118). Let $k = EPV/VHM$ and let $Z = \frac{n}{n+k} = \frac{n(VHM)}{n(VHM) + EPV}$. Then for **Bühlmann credibility**, (a linear approximation to Bayesian credibility), the credibility premium or Bühlmann credibility estimate is $P_C = Z\overline{X} + (1-Z)E(X) = E(X) + Z[\overline{X} - E(X)]$.

125) distribution	$1+[CV(X)]^2$
exponential EXP(θ)	2
gamma $G(\alpha,\theta)$	$1+\frac{1}{\alpha}$
lognormal $LN(\mu,\sigma)$	e^{σ^2}
Pareto(α,θ)	$\frac{2(\alpha-1)}{\alpha-2}$

126) For classical limited fluctuations credibility, $P_c = M + Z(T-M)$ where T is a statistic like $T = \overline{X}$ or $T = \sum_{i=1}^n X_i$ = total loss, and M is the prior estimate or manual rate or pure premium for what T estimates. Suppose you have $e < e_F, n < n_F$ or $a < a_F$ (expected) exposure units. Then $Z = \sqrt{e/e_F}, Z = \sqrt{n/n_F}$, or $Z = \sqrt{a/a_F}$.

127) The Bühlmann credibility method is a linear approximation to the Bayesian credibility method. Let $\Theta = \theta_i$ correspond to risk class i. Let the hypothetical mean $\mu_i = E(X|\Theta = \theta_i)$ be the mean of class i. The model or process is $X|\Theta$ (with pdf or pmf equal to the likelihood with $n=1$). Let $\mu = E(X) = E(E[X|\Theta]) = EHM$ = overall mean = expected value of the (process mean or) hypothetical mean. Let $v = EPV = E(V[X|\Theta])$ = expected value of the process variance. Let $a = VHM = V(E[X|\Theta])$ = variance of the hypothetical mean. Note that $V(X) = EPV + VHM$. Then Bühlmann's $k = \frac{v}{a} = \frac{EPV}{VHM}$ and Bühlmann's $Z = \frac{n}{n+k} = \frac{na}{na+v} = \frac{n(VHM)}{n(VHM)+EPV}$. Then for Bühlmann credibility, the credibility premium or Bühlmann credibility estimate is $P_c = Z\overline{X} + (1-Z)E(X) = E(X) + Z[\overline{X} - E(X)]$.

128) **Know:** For nonparametric or semiparametric empirical Bayes estimation for Bühlmann credibility, if $\hat{a} < 0$ set $\hat{a} = 0$ and $\hat{Z} = 0$.

129) For uniform exposures, nonparametric empirical Bayes estimation for Bühlmann credibility has X_{ij} = loss for the ith policy holder in the jth year, $\hat{\mu} = \overline{X} = \frac{1}{r}\sum_{i=1}^r \overline{X}_i = \frac{1}{nr}\sum_{i=1}^r\sum_{j=1}^n X_{ij}$, $\quad \hat{v} = \widehat{EPV} = \frac{1}{r}\sum_{i=1}^r \frac{1}{n-1}\sum_{j=1}^n (X_{ij} - \overline{X}_i)^2 = \frac{1}{r}\sum_{i=1}^r \hat{\sigma}^2_{ui}$, and
$\hat{a} = \widehat{VHM} = \frac{1}{r-1}\sum_{i=1}^r (\overline{X}_i - \overline{X})^2 - \frac{\hat{v}}{n}$. Here r = number of policyholders and n = number of years for loss data for each policyholder. If $\hat{a} = 0$ set $\hat{Z} = 0$, otherwise, $\hat{k} = \frac{\hat{v}}{\hat{a}}$ and $\hat{Z} = \frac{n}{n+\hat{k}}$. Then the Bühlmann premium for policyholder i is
$P_{ci} = \hat{Z}\overline{X}_i + (1-\hat{Z})\overline{X} = \overline{X} + \hat{Z}[\overline{X}_i - \overline{X}]$.

130) Nonuniform exposures, nonparametric empirical Bayes estimation for Bühlmann credibility: suppose there are n_i years of data for group (policyholder) i with m_{ij} exposures for group i in year j (uniform exposures 129) has $n_i \equiv n$ and $m_{ij} \equiv m$), and $m_i = \sum_{j=1}^{n_i} m_{ij}$. (If the time unit is years, then m_i is the number of exposure–years for group i over all n_i years.) Let $m = \sum_{i=1}^r m_i$. Then $\hat{\mu} = \overline{X} = \frac{\sum_{i=1}^r \sum_{j=1}^{n_i} m_{ij}X_{ij}}{m}$,

$\hat{v} = \widehat{EPV} = \dfrac{\sum_{i=1}^r \sum_{j=1}^{n_i} m_{ij}(X_{ij} - \overline{X}_i)^2}{\sum_{i=1}^r (n_i - 1)}$, and
$\hat{a} = \widehat{VHM} = \dfrac{\sum_{i=1}^r m_i(\overline{X}_i - \overline{X})^2 \quad - \quad \hat{v}(r-1)}{m - \frac{1}{m}\sum_{i=1}^r m_i^2}$. If $\hat{a} = 0$ set $\hat{Z} = 0$, otherwise, $\hat{k} = \dfrac{\hat{v}}{\hat{a}}$ and $\hat{Z} = \dfrac{n}{n + \hat{k}}$. Then the Bühlmann premium for policyholder i is
$P_c^i = P_{ci} = \hat{Z}\overline{X}_i + (1 - \hat{Z})\overline{X} = \overline{X} + \hat{Z}[\overline{X}_i - \overline{X}]$.

131) Semiparametric empirical Bayes estimation for Bühlmann credibility: assume the number of claims for each policyholder has a conditional Poisson(λ) distribution (λ is a RV). Each member has $\tilde{n} = 1$ year of exposure. The loss for the ith policy holder is X_i for $i = 1, ..., n$. (Note that n was r and $\tilde{n}$ was n in 129). Also note that $\overline{X}_i = X_i$ since $\tilde{n} = 1$.) Then $\hat{\mu} = \overline{X} = \hat{v} = \widehat{EPV}$, and $\hat{a} = \widehat{VHM} = \dfrac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 \quad - \quad \hat{v}$. If $\hat{a} = 0$ set $\hat{Z} = 0$, otherwise, $\hat{k} = \dfrac{\hat{v}}{\hat{a}}$ and $\hat{Z} = \dfrac{\tilde{n}}{\tilde{n} + \hat{k}} = \dfrac{1}{1 + \hat{k}}$. Then the Bühlmann premium for policyholder i is $P_c^i = P_{ci} = \hat{Z}X_i + (1 - \hat{Z})\overline{X} = \overline{X} + \hat{Z}[X_i - \overline{X}]$. This formula does not make sense with grouped data where 129) or 130) should be used.

132) Bayesian credibility: a) **Poisson–Gamma:** Suppose $N|\lambda \sim$ Poisson(λ) with conjugate prior distribution $\lambda \sim G(\alpha, \theta)$. Then $N \sim NB(r = \alpha, \beta = \theta)$. With k claims in n exposures, the posterior distribution $\lambda|(n,k) \sim G\left(\alpha' = \alpha + k, \theta' = \dfrac{\theta}{1 + n\theta}\right)$ where $\dfrac{1}{\theta'} = \dfrac{1}{\theta} + n = \dfrac{1 + n\theta}{\theta}$. The n exposures could be n years for one insured, n insureds for 1 year, or the sum of n_i insureds for year i for years 1, ..., m: $n = \sum_{i=1}^m n_i$. Note that $\overline{X} = k/n$. Then the posterior mean $E(\lambda|(n,k)) = \alpha'\theta' = P_c = \dfrac{(\alpha + k)\theta}{1 + n\theta} = \dfrac{\alpha + n\overline{X}}{\frac{1}{\theta} + n} = \dfrac{\gamma}{\gamma + n}\dfrac{\alpha}{\gamma} + \dfrac{n}{n + \gamma}\overline{X}$ where $\gamma = 1/\theta$ and $Z = \dfrac{n}{n + \gamma}$. The predictive distribution $N|(n,k) \sim NB\left(r = \alpha' = \alpha + k, \beta = \theta' = \dfrac{\theta}{1 + n\theta}\right)$.

b) normal–normal. Let $v = \sigma^2$ and $a = \tau^2$. Suppose $X \sim N(\theta, \sigma^2)$ with conjugate prior $\theta \sim N(\mu, \tau^2)$ where σ^2, μ and τ^2 are constants. (Sometimes use Θ in place of θ.) Then $X \sim N(\mu, \sigma^2 + \tau^2)$. Let the data $\boldsymbol{x} = (X_1, ..., X_n)$. Then the posterior distribution $\theta|\boldsymbol{x} \sim N(\mu + Z(\overline{X} - \mu), (1 - Z)\tau^2 + \sigma^2)$ with $Z = \dfrac{n}{n + \frac{\sigma^2}{\tau^2}} = \dfrac{n\tau^2}{n\tau^2 + \sigma^2}$.

c) binomial–beta: Suppose $N|q \sim$ binomial(q, m) with conjugate prior distribution $q \sim$ beta(a, b) Suppose there are k claims in m exposures. Then the posterior distribution $q|(m,k) \sim$ beta($a + k, b + m - k$). Here $q = P(claim)$.

133) If $N|q \sim$ Bernoulli(q) $\sim$ binomial($q, m = 1$) and $q \sim$ beta(a, b), and if the data $\boldsymbol{x}$ is n Bernoulli(q) trials with k 1's, then the posterior distribution $q|(n,k) \sim$ beta($a + k, b + n - k$). (132 c) treats the binomial(q,m) case as $n = m$ Bernoulli trials.) Then the posterior mean $E(q|\boldsymbol{x}) = \frac{a+k}{n+a+b}$.

134) For Bayesian credibility, typical exam questions tend to use distributions that are "easy to integrate" (to find the constant c that makes the posterior pdf integrate to

1) like the uniform, exponential, and single parameter Pareto distributions.

135) *Bernoulli shortcut:* Suppose W is a RV that takes on two values a and b with $p_a = P(W = a)$ and $p_b = 1 - p_a$. Then $V(W) = (b-a)^2 p_a p_b = (a-b)^2 p_a(1-p_a)$. $W = X|\Theta = \theta_i$ is possible. Note that $E(W) = ap_a + bp_b$.

136) Bühlmann credibility with a discrete prior $\pi(\text{class } i) = \pi_i$ where $\Theta = i$ denotes class i. Let X be the RV of interest. Let $\mu(\Theta) = E(X|\Theta)$ and $v(\Theta) = V(X|\Theta)$. Then $\mu(\theta) = E(X|\Theta = \theta)$ and $v(\theta) = V(X|\Theta = \theta)$. The model or process is $X|\Theta$.

row	class 1	...	class k	sum
i) prior	π_1	...	π_k	1
ii) $E(X\|$ class $i)$	μ_1	...	μ_k	
iii) $W_i = V(X\|$ class $i)$	$V(X\|1)$	...	$V(X\|k)$	

Then $\mu = E(\mu(\Theta)) = E(E(X|\Theta)) = E(X) = \sum_{i=1}^k \mu_i \pi_i$,
$v = E(v(\Theta)) = E(V(X|\Theta)) = \sum_{i=1}^k V(X|i)\pi_i = EPV$, and
$a = VHM = V(E(X|\Theta)) = V(W) = E(W^2) - [E(W)]^2 = E(W^2) - \mu^2 =$
$(\sum_{i=1}^k [\mu_i]^2 \pi_i) - \mu^2$.

Let $X = X_j$ for class j. Often k tables with n_j values of x_{ij} are given for $i = 1, ..., n_j$ and $j = 1, ..., k$ where the x_{ij} are the values X_j can take.

x_{i1}	$P(X_1 = x_{i1})$	...	x_{ik}	$P(X_k = x_{ik})$
x_{11}	p_{11}	...	x_{1k}	$p_{1,k}$
$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
$x_{n_1,1}$	$p_{n_1,1}$	...	$x_{n_k,k}$	$p_{n_k,k}$

Then $E(X_j) = \mu_j = \sum_{i=1}^{n_j} x_{ij} p_{ij}$ and $V(X_j) = V(X|j) = \sum_{i=1}^{n_j} (x_{ij} - \mu_j)^2 p_{ij}$. Often $k = 2$. Then $a = [\mu_2 - \mu_1]^2 \pi_1 \pi_2$ by 135). If $n_j = 2$, then $V(X|j) = (x_{2j} - x_{1j})^2 p_{1j} p_{2j}$ by 135).

137) If the Bühlmann premium for 1 member is $P_C^1 = E(X) + Z(\overline{X} - E(X))$, and the group has J members, then the Bühlmann premium for the group is $P_C = JP_C^1$.

138) In calculating Bühlmann's $Z = \dfrac{n}{n+k} = \dfrac{na}{na+v}$, need to know the number n of exposures. The exposure unit is the unit for which the credibility premium is charged. If you calculate the number of claims per insured, then the insured (a member) is the exposure unit. If you calculate claim size per claim then the exposure unit is a claim. Often an exposure unit is 1 member–year (member per year time period), and $n = n_1 + n_2 + \cdots + n_d$ where n_i is the number of members for the ith year, $i = 1, ..., d$.

139) The RV X for which you are calculating the credibility is often the claim count or claim size or aggregate loss of a single member of J members receiving insurance.

140) Bühlmann credibility with a continuous prior: the model or process is $X|\Theta$, the hypothetical mean is $\mu(\Theta) = E(X|\Theta)$, the process variance is $v(\Theta) = V(X|\Theta)$. Typically the prior is a brand name continuous distribution. Find $\mu = E(\mu(\Theta)) = E(E(X|\Theta)) = E(X)$, $v = E(v(\Theta)) = E(V(X|\Theta)) = EPV$, and $a = VHM = V(E(X|\Theta))$.

141) Often $X|\Lambda = N|\Lambda \sim$ Poisson(Λ). Then the hypothetical mean $\mu(\Lambda) = E(N|\Lambda) = \Lambda = V(N|\Lambda) = v(\Lambda) =$ the process variance. Then $a = V(\mu(\Lambda)) = V(\Lambda)$ and $v = E(v(\Lambda)) = E(\Lambda) = \mu$. See 142) = Table 51.1.

Material on Final but not on Exam 3.

143) Bühlmann Straub credibility: There are m_j exposures in period j. Assume $X_1, ..., X_n$ are independent conditional on Θ. Often $X_j = \overline{W}_j = \frac{1}{m_j}\sum_{i=1}^{m_j} W_{ij}$ where, conditional on $\Theta = \theta$, the W_{ij} are independent with mean $\mu(\theta) = E(X_j|\Theta = \theta)$ and variance $v(\theta)$ where $\dfrac{v(\theta)}{m_j} = V(X_j|\Theta = \theta)$. Then $\mu = E[\mu(\Theta)] = E[E(X_j|\Theta)] = E(X_j)$, $v = E[v(\Theta)] = E[V(X_j|\Theta)]$, and $a = V(\mu(\Theta)) = V(E(X_j|\Theta))$. Also $cov(X_i, X_j) = a$ for $i \neq j$ and $V(X_j) = \dfrac{v}{m_j} + a$, $k = v/a$, $Z = \dfrac{m}{m+k}$ where $m = \sum_{i=1}^n m_i$. Then $P_C^1 = Z\overline{X} + (1-Z)\mu = \mu + Z(\overline{X} - \mu)$ is the premium for 1 member of the group where $\overline{X} = \displaystyle\sum_{j=1}^n \frac{m_j}{m} X_j$. The credibility premium charged to the group in year $n+1$ is $P_C = m_{n+1}P_C^1$ where m_{n+1} is the number of group members in year $n+1$. Note that Bühlmann credibility has $m_j \equiv 1$ for $j = 1, ..., n$. Note that m_jX_j is the total loss for the group in year (time period) j: think of W_{ij} as the loss to the ith member in year j where there are m_j members in the group in year j. Then X_j is the average loss (of the m_j members) in year j.

144) **Inversion Method for a pdf** $X_i = F^{-1}(U_i)$: Let X be from a distribution with increasing cdf $F(x)$. Let $u_1, ..., u_n$ be pseudo U(0,1) random numbers. Then $x_1 = F^{-1}(u_1) = VaR_{u_1}(X), ..., x_n = F^{-1}(u_n) = VaR_{u_n}(X)$ are pseudo random numbers from the distribution of X. So $x_i = VaR_{u_i}(X)$ where $VaR_p(X)$ is given for several brand name distributions. If $VaR_p(X)$ is not given, solve $u = F(x)$ for $x = F^{-1}(u)$ and use $x_i = F^{-1}(u_i)$. Sometimes need to get the cdf $F(X) = \int_0^x f(t)dt$ where $f(t)$ is the pdf of a RV X with support $x > 0$.

145) **Inversion Method for a pmf:** Suppose the pmf has support $0, 1, ..., d, ..., J$ where $J = \infty$ is possible. Let $u_{(n)}$ be the largest U(0,1) pseudo number where $F(d-1) < u_{(n)} \leq F(d)$. Given $u_1, ..., u_n$, set $x_i = j$ if $F(j-1) \leq u_i < F(j)$ where $F(-1) = 0$, so set $x_i = 0$ if $0 \leq u_i < F(0)$. See the table below.

k	$p_k = P(X = k)$	$F(k)$	range of u	resulting x_i
0	p_0	$p_0 = F(0)$	$0 \leq u < F(0)$	0
1	p_1	$p_1 + F(0) = F(1)$	$F(0) \leq u < F(1)$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	p_j	$p_j + F(j-1) = F(j)$	$F(j-1) \leq u < F(j)$	j
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
d	p_d	$p_d + F(d-1) = F(d)$	$F(d-1) \leq u < F(d)$	d

146) Let $X_1, ..., X_n$ be iid random variables from a distribution with cdf F, mean μ and variance σ^2. Let $x_1, ..., x_n$ be the observed values of the X_i. The distribution of the RV D is the *empirical distribution* if D is a discrete RV with the following pmf.

x	x_1	x_2	$\cdots$	x_n
$P(D = x)$	$1/n$	$1/n$	$\cdots$	$1/n$

Then $E(D) = \overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^n x_i$ and $V(D) = \hat{\sigma}_E^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^n (x_i - \overline{x})^2$. If the x_i are not distinct, then let k_j = number of $x_i = x_j$, then $P(D = x_j) = k/n$, but this just combines columns in the above table that have $x_i = x_j$. The cdf of D is the empirical cdf F_n.

147) Let $X_1, ..., X_n$ be iid from a distribution with cdf F. Let $\theta = \theta(F)$. Let $T = g(x_1, ..., x_n)$ be an estimator of $\theta(F)$. Then $MSE_T(\theta(F)) = E[(T - \theta(F))^2] = V_\theta(T) + [\text{bias}_\text{T}(\theta(\text{F}))]^2$ where $\text{bias}_\text{T}(\theta(\text{F})) = \text{E}_\theta(\text{T}) - \theta(\text{F})$.

148) Refer to 147). The **bootstrap** uses B bootstrap samples where a bootstrap sample is a sample of size n drawn with replacement from $x_1, ..., x_n$ (iid wrt the empirical distribution). Let $x^*_{1i}, ..., x^*_{ni}$ denote the ith bootstrap sample. Let $T^*_i = g(x^*_{1i}, ..., x^*_{ni})$ be the statistic computed from $x^*_{1i}, ..., x^*_{ni}$ for $i = 1, ..., B$. Then the bootstrap approximation is $\widehat{MSE}_T(\theta(F)) = \widehat{MSE} = \frac{1}{B}\sum_{i=1}^{B}(T^*_i - T)^2$.

To use this approximation, compute the statistic T (eg the sample mean) from the sample and compute the statistic T^*_i (eg the sample mean) from the ith bootstrap sample for $i = 1, ..., B$.

149) Let $T = T_1$. If you had an iid sample $T_1, ..., T_B$ you could figure out how the statistic behaves, but you only have $T = T_1$. Under regularity conditions, if $\sqrt{n}(T_1 - \theta) \xrightarrow{D} N(\mu, \sigma^2)$, then $\sqrt{n}(T^*_i - T_1) \xrightarrow{D} N(\mu, \sigma^2)$. So $T^*_1 - T_1, ..., T^*_B - T_1$ is pseudodata for $T_1 - \theta, ..., T_B - \theta$.

150) Suppose you can simulate $\hat{\theta}$ and the true value of θ is known. If the n values of $\hat{\theta}_i$ are equally likely, then $\widehat{bias}(\hat{\theta}) = (\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i) - \theta$ and $\widehat{MSE}(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta)^2$.

151) $MSE(\overline{X}) = \dfrac{\sigma^2}{n}$. So $\widehat{MSE}(\overline{X}) = \dfrac{\hat{\sigma}^2}{n}$ where $\hat{\sigma}^2 = \hat{\sigma}^2_U$ or $\hat{\sigma}^2 = \hat{\sigma}^2_E$.

152) For the sample size n needed for simulation, see the back of HW 11, where $z_p = z_\pi$.