MAY 2007 SOA EXAM C/CAS 4 SOLUTIONS

1. Since there is left truncation at 50, each uncensored loss x contributes to the likelihood function a conditional density factor $f(x|X>50)=\frac{f(x)}{S(50)}=\frac{\frac{1}{\theta}e^{-x/\theta}}{e^{-50/\theta}}=\frac{1}{\theta}e^{-(x-50)/\theta}$,
and each right-censored loss x contributes a conditional survival factor
$P(X>x|X>50)=\frac{S(x)}{S(50)}=\frac{e^{-x/\theta}}{e^{-50/\theta}}=e^{-(x-50)/\theta}$.
Note that x refers to the actual censored loss amount before application of the deductible of 50. We are given claim payments, so the actual loss amount x is claim payment $+\,50$ for the uncensored claim payments, and the right censored loss amount are 400 (censored claim payment of 350 plus 50). The uncensored losses are 100 , 200 , 250 , and there are two censored loss of 400. The likelihood function is
$L(\theta)=\frac{1}{\theta}e^{-(100-50)/\theta}\cdot\frac{1}{\theta}e^{-(200-50)/\theta}\cdot\frac{1}{\theta}e^{-(250-50)/\theta}\cdot e^{-(400-50)/\theta}\cdot e^{-(400-50)/\theta}$
$=\frac{1}{\theta^3}e^{-(1100)/\theta}$. Answer: D

2. The Buhlmann estimate is $ZX+(1-Z)\mu$, where X is the number of claims in the first year. This is a linear function of X, so answer D can be eliminated, since the graph of the Buhlmann credibility premium in D is not a linear function of X.

The Bayesian estimate is
$E[X_2|X_1=k]=\int_{.1}^{.6}E[X_2|q]\cdot\pi(q|X_1=k)\,dq=\int_{.1}^{.6}6q]\cdot\pi(q|X_1=k)\,dq=6E[q|X_1=k]$.
Since $.1\le q\le .6$. we must have $.1\le E[q|X_1=k]\le .6$, and therefore the Bayesian estimate must be between .6 and 3.6. This eliminates answers A and C, since A has a Bayesian estimate greater than 3.6 and C has a Bayesian estimate less than .6 (and one greater than 3.6).

The following reasoning can be used to eliminate answer B. The Buhlmann premium is $ZX+(1-Z)E[X]$. The expected Buhlmann premium (expected based on the value of X) is $ZE[X]+(1-Z)E[X]=E[X]$ (the mean of the marginal distribution of X).
The Bayesian premium is $E[X_2|X_1]$. From the double expectation rule, we know that $E[\,E[X_2|X_1]\,]=E[X_2]$, which is the mean of the marginal distribution of X.
In graph B, the Bayesian premium is less than the Buhlmann premium for every value of X, so the expected value of the Bayesian premium would be less than that of the Buhlmann premium. Since this is not possible, graph B cannot be correct. Answer: E

3. $Var[S_{101}]=Var[\,E[S_{101}|Q]\,]+E[\,Var[S_{101}|Q]\,]$.
The conditional distribution of S_{101} given Q is binomial with $m=101$ and Q.
$E[S_{101}|Q]=101Q$ and $Var[S_{101}|Q]=101Q(1-Q)$.
From the Exam C table of distributions, we have
$E[Q]=\frac{a}{a+b}=\frac{1}{1+99}=\frac{1}{100}$ and $E[Q^2]=\frac{a(a+1)}{(a+b)(a+b+1)}=\frac{2}{(100)(101)}$.
Then, $Var[\,E[S_{101}|Q]\,]=Var[101Q]=101^2Var[Q]=101^2[\frac{2}{(100)(101)}-\frac{1}{100^2}]=.9999$
and
$E[\,Var[S_{101}|Q]\,]=E[101Q(1-Q)]=101(E[Q]-E[Q^2])=101(\frac{1}{100}-\frac{2}{(100)(101)})=.99$.
Therefore, $Var[S_{101}]=.9999+.99=1.9899$. Answer: B

4. The lognormal model for the price of the stock at time t is
$ln(\frac{S_t}{S_0})$ has a normal distribution with mean $(\alpha-\delta-\frac{1}{2}\sigma^2)t$ and variance $\sigma^2 t$.
We are given that the current price is $S_0=0.25$, $\alpha=0.15$, $\delta=0$ and $\sigma=0.35$,
so that $ln(\frac{S_5}{S_0})$ has a normal distribution with mean $.044375$ and variance $.06125$.
The upper bound of the 90% confidence interval would be the 95-th percentile of the stock price in 6 months, say c. Then
$P(S_{.5}\le c)=P(\frac{S_{.5}}{S_0}\le\frac{c}{.25})=P[ln(\frac{S_{.5}}{S_0})\le ln(\frac{c}{.25})]=\Phi[\frac{ln(\frac{c}{.25})-.044375}{\sqrt{.06125}}]=.95$.
It follows that $\frac{ln(\frac{c}{.25})-.044375}{\sqrt{.06125}}=1.645$, from which we get $c=.3927$. Answer: A

5. Since the 20 ranges are of equal length, each length is .05 (one-twentieth of 1), and we expect $E_i=50$ observations in each subinterval (one-twentieth of 1000). The Chi-square test statistic is
$Q=\sum_{i=1}^{20}\frac{(O_i-E_i)^2}{E_i}=\sum_{i=1}^{20}\frac{(O_i-50)^2}{50}$.
$\sum_{i=1}^{20}(O_i-50)^2=\sum_{i=1}^{20}O_i^2-2(50)\sum_{i=1}^{20}O_i+20(50^2)$.
Since there are 1000 pseudo-random numbers, we have $\sum_{i=1}^{20}O_i=1000$, so that
$\sum_{i=1}^{20}(O_i-50)^2=51,850-2(50)(1000)+20(50^2)=1850$.
Then, $Q=\frac{1850}{50}=37$. The degrees of freedom in the test is $20-1=19$, since no model parameters are estimated. From the Exam C tables, we see that for the Chi-square distribution with 19 degrees of freedom, the 99-th percentile is 36.191 and the 99.5-th percentile is 38.582. We reject H_0 at the 1% level but not at the .5% level because $36.191<37<38.582$.
Answer: E

6. The credibility factor is $Z=\frac{n}{n+\frac{v}{a}}$. In this case, $n=1$.
The hypothetical means are $E[X|I]=.5$ with probability (proportion) θ, and $E[X|II]=1.5$ with probability (proportion) $1-\theta$. The process variances are $Var[X|I]=.5$ with probability (proportion) θ, and $Var[X|II]=1.5$ with probability (proportion) $1-\theta$.

The variance of the hypothetical mean is the variance of the two-point random variable
$E[X|I]=.5$, prob. θ , $E[X|II]=1.5$, prob. $1-\theta$.
This variance is $a=(.5-1.5)^2\theta(1-\theta)=\theta-\theta^2$.

The expected process variance is the expected value of the two-point random variable
$Var[X|I]=.5$, prob. θ , $Var[X|II]=1.5$, prob. $1-\theta$. This expected value is
$v=.5\theta+1.5(1-\theta)=1.5-\theta$.

The credibility factor is $Z=\frac{1}{1+\frac{1.5-\theta}{\theta-\theta^2}}=\frac{\theta-\theta^2}{1.5-\theta^2}$. Answer: A

7. The limit of 350 is in the interval $(200, 400]$, which has 16 claims. Applying the uniform distribution to that interval, there would be $(.75)(16) = 12$ claims in the interval $(200, 350]$ and 4 claims larger than 350. Then applying the 2nd moment equation to each interval, and the limit point, the estimate for $E[(X \wedge 350)^2]$ is
$\frac{30}{100} \cdot \frac{50^3-0^3}{3(50-0)} + \frac{36}{100} \cdot \frac{100^3-50^3}{3(100-50)} + \frac{18}{100} \cdot \frac{200^3-100^3}{3(200-100)} + \frac{12}{100} \cdot \frac{350^3-200^3}{3(350-200)} + \frac{4}{100} \cdot 350^2$
$= 20,750$. Answer: E

8. S has a compound Poisson distribution.
$P(S \leq 3) = P(S = 0) + P(S = 1) + P(S = 2) + P(S = 3)$.

$P(S = 0) = P(N = 0) = e^{-3}$.

$P(S = 1) = P(N = 1) \cdot P(X = 1) = \frac{e^{-3}3^1}{1!} \cdot (.4) = 1.2e^{-3}$.

$P(S = 2) = P(N = 1) \cdot P(X = 2) + P(N = 2) \cdot [P(X = 1)]^2$
$= \frac{e^{-3}3^1}{1!} \cdot (.3) + \frac{e^{-3}3^2}{2!} \cdot (.4)^2 = 1.62e^{-3}$.

$P(S = 3) = P(N = 1) \cdot P(X = 3) + P(N = 2) \cdot 2P(X = 1)P(X = 2)$
$+ P(N = 3) \cdot [P(X = 1)]^3$
$= \frac{e^{-3}3^1}{1!} \cdot (.2) + \frac{e^{-3}3^2}{2!} \cdot 2(.4)(.3) + \frac{e^{-3}3^3}{3!} \cdot (.4)^3 = 1.968e^{-3}$.
Aggregate losses do not exceed 3 if $S \leq 3$.
$P(S \leq 3) = e^{-3}[1 + 1.2 + 1.62 + 1.968] = .288$. Answer: B

9. For frequency, N, we have

n	0	1	2	3	4	5	...
$F(n)$	.5556	.8025	.9122	.9610	.9827	.9923	...

We are given .981 to simulate n. Since $F(3) = .9610 \leq .981 < .9827 = F(4)$, the simulated value of n is 4.

The distribution function of the Pareto severity is
$F(x) = 1 - \left(\frac{\theta}{x+\theta}\right)^\alpha = 1 - \left(\frac{36}{x+36}\right)^{2.8}$.
We must simulate 4 Pareto severity claim amounts, and then apply the deductible and maximum payment. We find the simulated claim amount x from the uniform random number u by solving for x from the equation $u = 1 - \left(\frac{36}{x+36}\right)^{2.8}$. This results in $x = 36(1-u)^{-1/2.8} - 36$.
For $u_1 = .571$ we get $x_1 = 12.7$, for $u_2 = .932$ we get $x_2 = 58.0$,
for $u_3 = .303$ we get $x_3 = 5.0$, and for $u_4 = .471$ we get $x_3 = 9.2$.
We apply the deductible to all claims, resulting values (before the maximum payment of 30 is imposed) of 7.7 , 53 , 0 , 4.2 . We impose the maximum payment of 30 to the second claim, resulting in total simulated claim payments of $7.7 + 30 + 4.2 = 41.9$. Answer: B

10. The mean of the shifted exponential is $\int_\delta^\infty x \cdot \frac{1}{\theta} e^{-(x-\delta)/\theta}\, dx$.
With the change of variable $y = x - \delta$, this integral becomes
$\int_0^\infty (y+\delta) \cdot \frac{1}{\theta} e^{-y/\theta}\, dy = \int_0^\infty y \cdot \frac{1}{\theta} e^{-y/\theta}\, dy + \delta \int_0^\infty \frac{1}{\theta} e^{-y/\theta}\, dy = \theta + \delta$
(the first integral is just the mean of an exponential with mean θ, and the second integral is δ times the pdf of the exponential).
The median of the shifted exponential is m, where $\int_\delta^m f(x)\, dx = .5$.
Since $\int_\delta^m \frac{1}{\theta} e^{-(x-\delta)/\theta}\, dx = 1 - e^{-(m-\delta)/\theta}$, we solve $1 - e^{-(m-\delta)/\theta} = .5$ for x
to get $m = \delta - \theta\, ln(.5)$.
We then set $\theta + \delta = 300$ (population mean = sample mean) , and
$\delta - \theta\, ln(.5) = 240$ (population median = sample median).
Subtracting the second equation from the first results in $\theta[1 + ln(.5)] = 60$, so that $\theta = 195.5$
and then $\delta = 300 - \theta = 104.5$. Answer: E

11. There are $r = 3$ policyholders and $n = 3$ observations for each policyholder.
The credibility factor is $Z = \frac{3}{3+\frac{v}{a}} = \frac{3}{3+\frac{v}{a}}$, where $v = \frac{1}{r(n-1)} \sum_{i=1}^{r} \sum_{j=1}^{n} (X_{ij} - \bar{X}_i)^2$
and $a = \frac{1}{r-1} \sum_{i=1}^{r} (\bar{X}_i - \bar{X})^2 - \frac{v}{n}$.
Since we are given the sample variance $\frac{1}{n-1} \sum_{j=1}^{n} (X_{ij} - \bar{X}_i)^2$ for each policyholder, we get
for policyholder I, $\sum_{j=1}^{3} (X_{Ij} - \bar{X}_I)^2 = 2(1) = 2,$
for policyholder II, $\sum_{j=1}^{3} (X_{IIj} - \bar{X}_{II})^2 = 2(3) = 6$,
and for policyholder III, $\sum_{j=1}^{3} (X_{IIIj} - \bar{X}_{III})^2 = 2(1) = 2.$
Then, $v = \frac{1}{3(3-1)} [2 + 6 + 2] = \frac{5}{3}$.
Since all policyholders have the same number of observed data points, we have
$\bar{X} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}{3} = \frac{5+9+6}{3} = \frac{20}{3}$.
Then $a = \frac{1}{3-1} [(5 - \frac{20}{3})^2 + (9 - \frac{20}{3})^2 + (6 - \frac{20}{3})^2] - \frac{5/3}{3} = \frac{34}{9}$.
Finally, $Z = \frac{3}{3 + \frac{5/3}{34/9}} = .872$. Answer: D

12. Since there is no censoring, the product limit estimator of $S(t)$ is $\frac{n_t}{200}$, where n_t is the number of survivors at time t. It follows that $n_8 = (200)(.22) = 44$, and $n_9 = 32$. Also, since there is no censoring, $n_8 = 44$ is the number at risk at "death" point $t = 9$, so $r_9 = n_8 = 44$, and similarly, $r_{10} = n_9 = 32$. Note that we can also see that the number of "deaths" at time $t = 9$ is $s_9 = 44 - 32 = 12$.

Greenwood's approximation is $Var[\widehat{S}(y_i)] = [\widehat{S}(y_i)]^2 \cdot \sum_{j=1}^{i} \frac{s_j}{r_j(r_j - s_j)}$,

so that $c_S^2(y_i) = \frac{V\hat{a}r[\widehat{S}(y_i)]}{\widehat{S}(y_i)^2} = \sum_{j=1}^{i} \frac{s_j}{r_j(r_j - s_j)}$.

Then, $c_S^2(9) = \sum_{j=1}^{9} \frac{s_j}{r_j(r_j - s_j)} = .02625$, and $c_S^2(10) = \sum_{j=1}^{10} \frac{s_j}{r_j(r_j - s_j)} = .04045$.

It follows that $c_S^2(10) - c_S^2(9) = \frac{s_{10}}{r_{10}(r_{10} - s_{10})} = .04045 - .02625 = .0142$.

Then, since $r_{10} = 32$, we have $\frac{s_{10}}{32(32 - s_{10})} = .0142$.

Solving for s_{10} results in $s_{10} = 10$. Answer: A

13. The coefficient of variation of Y is $\frac{\sqrt{Var(Y)}}{E(Y)}$.
$Y = X - (X \wedge 30000)$, so $E(Y) = E(X) - E(X \wedge 30000)$.
For an exponential random variable W with mean θ, $E(W \wedge u) = \theta(1 - e^{-u/\theta})$.
It follows that $E(X \wedge 30000) = 10,000(1 - e^{-30,000/10,000}) = 9502$.
Then, $E(Y) = 10,000 - 9502 = 498$.
$Var(Y) = E(Y^2) - [E(Y)]^2$.
$E(Y^2) = \int_{30000}^{\infty} (x - 30000)^2 \cdot \frac{1}{10,000} e^{-x/10,000}\, dx$.
Using the change of variable $z = x - 30,000$, this integral becomes
$\int_0^{\infty} z^2 \cdot \frac{1}{10,000} e^{-(z+30,000)/10,000}\, dz = e^{-3} \cdot \int_0^{\infty} z^2 \cdot \frac{1}{10,000} e^{-z/10,000}\, dz$
$= e^{-3} \cdot (2 \times 10,000^2) = 9,957,414$ (the integral is the 2nd moment of an exponential random variable with mean 10,000).
Then, $Var(Y) = 9,957,414 - (498)^2 = 9,709,410$.
The coefficient of variation of Y is $\frac{\sqrt{9,709,410}}{498} = 6.26$.
Note that we can use the following approach to get $E(Y^2)$.
If W has an exponential distribution with mean θ, then he conditional distribution of $W - d$ given that $W > d$ also has an exponential distribution with mean θ.
Therefore, $E[(W - d)^2 | W > d] = 2\theta^2$. But it is also true that
$E[(W - d)^2 | W > d] = \frac{E[(W-d)_+^2]}{P(W>d)}$, so that
$E[(W - d)_+^2] = E[(W - d)^2 | W > d] \cdot P(W > d) = (2\theta^2)(e^{-d/\theta})$.
Applying this to X, and $Y = (X - 30000)_+$, we see that
$E(Y^2) = E[(X - 30000)_+^2] = E[(X - 30000)^2 | X > 30000] \cdot P(X > 30000)$
$= (2 \times 10,000^2)(e^{-30,000/10,000}) = 9,957,414$, as before. Answer: C

14. The pdf of the Pareto distribution is $f(x) = \frac{\alpha\theta^\alpha}{(x+\theta)^{\alpha+1}}$, for $x > 0$.
The natural log of the pdf is $ln\, f(x) = ln\, \alpha + (\alpha+1)\, ln\, \theta - \alpha\, ln(x+\theta)$.
The log of the likelihood is the sum of the log of the pdf for each of the data points, so if there are n data points,
$ln\, L = \Sigma ln[f(x_i)] = n\, ln\, \alpha + n\alpha\, ln\, \theta - (\alpha+1)\Sigma\, ln(x_i+\theta)$.

Since $n = 20$, the loglikelihood for the hypothesized distribution (Pareto with $\alpha = 2$ and $\theta = 3.1$) is
$ln\, L_0 = 20\, ln(2) + 20(2)\, ln(3.1) - 3\Sigma\, ln(x_i + 3.1)$
$= 13.863 + 45.256 - 3(39.30) = -58.781$.

The loglikelihood for the alternative distribution (Pareto with $\alpha = 2$ and $\theta = 7.0$) is
$ln\, L_1 = 20\, ln(2) + 20(2)\, ln(7.0) - 3\Sigma\, ln(x_i + 7.0)$
$= 13.863 + 77.836 - 3(49.01) = -55.331$.

Only the parameter θ is estimated. The degrees of freedom in the test is 1 (the hypothesized distribution has no parameters estimated, and the alternative distribution has 1 parameter estimated). The test statistic for the Chi-square test is
$2(ln\, L_1 - ln\, L_0) = 2(-55.331 + 58.781) = 6.9$.
In the Chi-square table with 1 degree of freedom, we see that the 99-th percentile is 6.636.
Since $6.9 > 6.636$, the null hypothesis is rejected at the 1% level, (it is rejected at he .5% level as well). Answer: E

15. We have the combination of the beta prior distribution for q and the binomial model distribution with parameters m and q. Given n observed values $x_1, x_2, ..., x_n$, the Bayesian credibility estimate is $m \cdot \frac{a'}{a'+b'}$, where $a' = a + \Sigma x_i$ and $b' = b + nm - \Sigma x_i$
(a' and b' are the parameters of the posterior distribution of q, which is also a beta distribution).
We are given that $m = 8$ in the binomial model distribution.

From (iv), we have $n = 1$, and $x_1 = 2$, so with prior values a (unknown) and $b = 9$,
the Bayesian credibility estimate for Year 2 is $m \cdot \frac{a'}{a'+b'} = 8 \cdot \frac{a+2}{(a+2)+[9+(1)(8)-2]} = 2.54545$.
We can solve this equation for a, resulting in $a = 5$.

Then, from (v), we have $n = 2$, and $x_1 = 2$, $x_2 = k$ (unknown), so with prior values $a = 5$ and $b = 9$, the Bayesian credibility estimate for Year 3 is
$m \cdot \frac{a'}{a'+b'} = 8 \cdot \frac{5+2+k}{(5+2+k)+[9+(2)(8)-2-k]} = 8 \cdot \frac{7+k}{30} = 3.73333$.
We can solve this equation for a, resulting in $k = 7$. Answer: D

16. The empirical distribution function is $F_5(150) = \frac{2}{5} = .4$.
With bandwidth $b = 50$ at $x = 150$, we have the following.
The full interval centered at $y_1 = 82$ (from 32 to 132) is to the left of $x = 150$,
so that $K_{82}(150) = 1$.
The interval centered at $y_2 = 126$ is from 76 to 176, so $K_{126}(150) = \frac{150-76}{176-76} = .74$.
The interval centered at $y_3 = 161$ is from 111 to 211, so $K_{161}(150) = \frac{150-111}{211-111} = .39$.
The interval centered at $y_4 = 294$ is from 244 to 344, and is completely to the right of $x = 150$,
so $K_{294}(150) = 0$. The same applies to the interval centered at $y_5 = 384$.
The kernel smoothed estimate of $F(150)$ is
$\widehat{F}(150) = \frac{1}{5} \cdot (1) + \frac{1}{5} \cdot (.74) + \frac{1}{5} \cdot (.39) = .426$.
Then, $|.426 - .4| = .026$. Answer: C

17. Aggregate losses will be represented by the random variable S.
We wish to find $P[S > 1.5E(S)]$, using the normal approximation.
This will be $P[\frac{S-E(S)}{\sqrt{Var(S)}} > \frac{1.5E(S)-E(S)}{\sqrt{Var(S)}}] = 1 - \Phi(\frac{.5E(S)}{\sqrt{Var(S)}})$.
For the compound distribution, we have
$E(S) = E(N) \cdot E(X) = (100)(20,000) = 2,000,000$
(N is the claim count, and X is the single-loss).
We also have
$Var(S) = E(N) \cdot Var(X) + Var(N) \cdot [E(X)]^2 = (100)(5000^2) + (25^2)(20,000^2)$
$= 2.525 \times 10^{11}$.
The probability is $1 - \Phi(\frac{.5E(S)}{\sqrt{Var(S)}}) = 1 - \Phi(\frac{.5(2,000,000)}{\sqrt{2.525 \times 10^{11}}}) = 1 - \Phi(1.99)$
$= 1 - .9761 = .0239$. Answer: A

18. The maximum likelihood estimate of a Poisson parameter is the sample mean.
This estimate will be $\widehat{\lambda} = \frac{1200(1)+600(2)+200(3)}{3000} = 1.00$.
Since X, the number of claims per policy, has a Poisson distribution, the estimated variance is also 1.00 . The estimator $\widehat{\lambda}$ is the sample mean of 3000 observations, so the variance of the estimator is $Var(\widehat{\lambda}) = Var(\overline{X}) = \frac{Var(X)}{3000}$. Using the estimate for the variance of X of 1, we get an estimated variance of $\widehat{\lambda}$ of $\frac{1}{3000}$. The 90% linear symmetric confidence interval for λ is
$\widehat{\lambda} \pm 1.645\sqrt{V\widehat{a}r(\widehat{\lambda})} = 1.00 \pm 1.645\sqrt{\frac{1}{3000}} = (.97\,,\ 1.03)$. Answer: C

19. The model for the stock price at time 2 is $S_2 = S_0 e^{(\alpha - \frac{\sigma^2}{2})(2) + \sigma\sqrt{2}Z}$,
where Z is a standard normal random variable. We use the given $(0,1)$ values to simulate standard normal values. The simulated standard normal values are found by the inversion method from the normal table. The simulated normal values are 2.12 , -1.77 , 0.77 .
The simulated stock prices are $S_0 e^{(\alpha - \frac{\sigma^2}{2})(2) + \sigma\sqrt{2}Z} = 50e^{(.15 - \frac{.09}{2})(2) + .3\sqrt{2}Z}$.
Substituting in the three simulated values for Z results in simulated prices of
151.63 , 29.11 , 85.52 . The mean of these three values is 88.75 . Answer: C

20. The maximum value of $|F_n(x) - F^*(x)|$ is $.111$ and occurs at $x = 4.39$.
With $n = 200$, the critical values are $\frac{1.22}{\sqrt{200}} = .0863$ at 10% significance,
$\frac{1.36}{\sqrt{200}} = .0962$ at 5% significance, $\frac{1.52}{\sqrt{200}} = .1075$ at 2% significance,
and $\frac{1.63}{\sqrt{200}} = .1153$ at 1% significance. Since $.1075 < .111 < .1153$,
the test result is to reject H_0 at the 2% level, but not at the 1% level. Answer: D

21. The Buhlmann credibility factor based on $n = 2$ observations is $Z = \frac{2}{2+\frac{v}{a}} = .25$
The hypothetical mean is $E(X|\alpha) = \alpha\theta$ (mean of the gamma distribution).
The process variance is $Var(X|\alpha) = \alpha\theta^2$ (variance of the gamma distribution).
The expected process variance is $v = E[Var(X|\alpha)] = E[\alpha\theta^2] = \theta^2 E(\alpha) = 50\theta^2$.
The variance of the hypothetical mean is $a = Var[E(X|\alpha)] = Var(\alpha\theta) = \theta^2 Var(\alpha)$.
From $Z = \frac{2}{2+\frac{v}{a}} = .25$ we get that $\frac{v}{a} = 6$, so that $\frac{50\theta^2}{\theta^2 Var(\alpha)} = 6$.
From this equation we see that $Var(\alpha) = \frac{50}{6}$. Answer: A

22. We first can observe that the baseline survival function is a single parameter Pareto survival with $\theta = 200$ and unknown α. The survival function for someone from Group 2 will have the form $S(x) = [S_0(x)]^c = \left(\frac{200}{x}\right)^{\alpha c}$, where $c = e^{\beta}$.; this is also a Pareto survival function, also with $\theta = 200$, but with αc instead of α. The pdf of the baseline single parameter Pareto is
$f_0(x) = \frac{\alpha\theta^{\alpha+1}}{x^{\alpha}} = \frac{\alpha \cdot 200^{\alpha+1}}{x^{\alpha}}$, and pdf of the Group 2 single parameter Pareto is
$f(x) = \frac{\alpha\theta^{\alpha+1}}{x^{\alpha}} = \frac{c\alpha \cdot 200^{c\alpha+1}}{x^{c\alpha}}$.
The natural log of baseline pdf is $ln\, f_0(x) = ln\,\alpha + (\alpha+1)\,ln\,200 - \alpha\,ln\,x$
and the natural log of Group 2 pdf is $ln\, f(x) = ln\, c + ln\,\alpha + (c\alpha+1)\,ln\,200 - c\alpha\,ln\,x$.
The loglikelihood is

$$ln\,L = 3\,ln\,\alpha + 3(\alpha+1)\,ln\,200 - \alpha\sum_{\text{baseline}} ln\,x_i$$

$$+ 3\,ln\,c + 3\,ln\,\alpha + 3(c\alpha+1)\,ln\,200 - c\alpha\sum_{\text{Group 2}} ln\,x_i$$

The partial derivative with respect to α is

$$\frac{\partial}{\partial\alpha} ln\,L = \frac{6}{\alpha} + 3(1+c)\,ln\,200 - \sum_{\text{baseline}} ln\,x_i - c\sum_{\text{Group 2}} ln\,x_i$$

$$= \frac{6}{\alpha} + 3(1+c)\,ln\,200 - 16.5959c - 17.6544 = \frac{6}{\alpha} - .7009c - 1.7593 \ .$$

The partial derivative with respect to c is

$$\frac{\partial}{\partial c} ln\,L = \frac{3}{c} + 3\alpha\,ln\,200 - \alpha\sum_{\text{Group 2}} ln\,x_i = \frac{3}{c} + 3\alpha\,ln\,200 - 16.5959\alpha \ .$$

$$= \frac{3}{c} - .7009\alpha \ .$$

We set both partial derivatives equal to 0.
From $\frac{3}{c} - .7009\alpha = 0$ we get $\alpha = \frac{4.280}{c}$.
Then, from $\frac{6}{\alpha} - .7009c - 1.7593 = 0$, we get $\frac{6c}{4.280} - .7009c - 1.7593 = 0$, from which we get $c = 2.51$. Then, since $c = e^{\beta}$, we get $\beta = ln(2.51) = .92$. Answer: D

23. Ruin can occur in the first year only if there is a loss of 6 in the first year. This probability is 0.1. Ruin can occur in the 2nd year if the loss is 0 or 2 in the first year and the loss is 6 in the second year. This probability is $.9 \times .1 = .09$.
Ruin can occur in the 3rd year if
(i) ruin has not yet occurred by the start of the 3rd year, and
(ii) if surplus is less than $\frac{6}{1.1} = 5.45$ after premium is received at the start of the 3rd year, and
(iii) the loss in the 3rd year is 6.
(i) and (ii) occur only if there is a loss of 2 in the first year and a loss of 2 in the second year.
(if the first two year losses are 0 and 0, then there will be surplus of 7.83 after premium is received at the start of the third year, if the first two year losses are 0 and 2 there will be surplus of 5.83, and if the first two year losses are 2 and 0 there will surplus of 5.63 at the start of the third year after premium is received, so ruin cannot occur in those cases).
The probability of ruin occurring in the third year is $(.3)(.3)(.1) = 009$.
The total probability of ruin in the first three years is $.1 + .09 + .009 = .199$. Answer: C

24. Since there $n = 16$ data points, we construct the smoothed empirical percentiles (SEP) as

x_i	54	...	75	81	...
SEP	$\frac{1}{17} = .0588$	...	$\frac{3}{17} = .1765$	$\frac{4}{17} = .2353$	...

and

x_i	...	122	125	...
SEP	...	$\frac{11}{17} = .6471$	$\frac{12}{17} = .7059$	...

The smoothed estimate of the 20th percentile is between 75 and 81.
The interpolation from $\frac{3}{17}$ to $.20$ to $\frac{4}{17}$ is the same as from 3 to 3.4 to 4 (multiply by 17), so the smoothed estimate of the 20th percentile is 40% of the way from 75 to 81, which is 77.4 . In a similar way, the smoothed estimate of the 70th percentile is between 122 and 125. The interpolation from $\frac{11}{17}$ to $.70$ to $\frac{12}{17}$ is the same as from 11 to 11.9 to 12 (multiply by 17), so the smoothed estimate of the 70th percentile is 90% of the way from 122 to 125, which is 124.7 .

The distribution function of the Weibull is $F(x) = 1 - e^{-(x/\theta)^\tau}$.
The two percentile equations are $F(77.4) = 1 - e^{-(77.4/\theta)^\tau} = .20$
and $F(124.7) = 1 - e^{-(124.7/\theta)^\tau} = .70$.
These equations become $\left(\frac{77.4}{\theta}\right)^\tau = -ln(.8)$ and $\left(\frac{124.7}{\theta}\right)^\tau = -ln(.3)$.
Dividing the 2nd equation by the first, we get $\left(\frac{124.7}{77.4}\right)^\tau = \frac{ln(.3)}{ln(.8)} = 5.3955$,
and then $\tau = \frac{ln(5.3955)}{ln(\frac{124.7}{77.4})} = 3.53$. Then from $\left(\frac{77.4}{\theta}\right)^\tau = \left(\frac{77.4}{\theta}\right)^{3.53} = -ln(.8)$
we get $\theta = 118$. Answer: E

25. The data we have is for a 5-ear period, so the empirical Bayes estimate will be the number of claims for the next 5 year period. We would then divide that estimate by 5 to get the estimate for the number of claims in Year 6. According to the semiparametric method, the estimate for the number of claims in the next 5 year period is $Z\overline{Y} + (1-Z)\mu$, where $\overline{Y}$ is the sample mean of the data points available for the selected policyholder. In this example, we have one observed value $Y = 3$ as the number of claims in the first 5 year period for the selected policyholder.

25 continued
There is 1 observations, so $\overline{Y} = 3$. The credibility factor is $Z = \frac{1}{1+\frac{v}{a}}$, since there is only one observed value for the selected policyholder. For this Poisson semiparametric model, the estimates of μ, v and a are $\widehat{\mu} = \widehat{v} = \overline{X}$, and $\widehat{a} = s_X^2 - \overline{X}$.
From the given data set, $\overline{X} = \frac{34+13(2)+(5(3)+2(4)}{100} = .83$, and
$s_X^2 = \frac{1}{99}[\Sigma X_i^2 - 100\overline{X}^2] = \frac{1}{99}[34 + 13(4) + 5(9) + 2(16) - 100(.83)^2] = .95$.
Then, $\widehat{\mu} = \widehat{v} = .83$ and $\widehat{a} = .95 - .83 = .12$.
The estimated credibility factor is $\widehat{Z} = \frac{1}{1+\frac{.83}{.12}} = .126$, and the estimate for the number of claims in the next 5 years for the selected policyholder is $(.126)(3) + (.874)(.83) = 1.1$.
The estimate for the number of claims in Year 6 for the selected policyholder is $\frac{1.1}{5} = .22$.
Answer: A

26. With a deductible of 500, the claim payment will be above 5500 if the loss amount is above 6000. We wish to estimate $P(X > 6000|X > 500) = \frac{S(6000)}{S(500)}$.
Using the large sample approach for product limit estimation, the estimate of $S(c_i)$ is
$\widehat{S}(c_i) = (1 - \frac{x_0}{r_0})\cdots(1 - \frac{x_i}{r_i})$. The data is organized in interval form, with intervals $(c_0, c_1]$, $(c_1, c_2]$, ... The first row of data tells us that $c_0 = 250$, $d_0 = 6$, etc.
The table gives us the x_j values, so in order to find the estimates, we need to find the r_j values.
With α and β, we have $r_0 = \alpha d_0 - \beta u_0$.
In this case that is $r_0 = d_0 = 6$. The general calculation for r_i is
$r_i = P_i - \alpha d_i - \beta u_i$. We are given $P_1 = 5$, $P_2 = 9$, ...
Using this, we get
$r_1 = 5 + 6 = 11$, $r_2 = 9 + 7 = 16$, $r_3 = 11$, $r_4 = 3$.
Since the data is initially truncated at 250, from the first interval right endpoint ($c_1 = 500$) we can estimate $P(X > 500|X > 250) = 1 - \frac{x_0}{r_0} = 1 - \frac{1}{6} = \frac{5}{6} = .8333.$
The estimate of $P(X > 6000|X > 250)$ (5th interval endpoint is $c_5 = 6000$)
$(1 - \frac{x_0}{r_0})(1 - \frac{x_1}{r_1})(1 - \frac{x_2}{r_2})(1 - \frac{x_3}{r_3})(1 - \frac{x_4}{r_4})$
$= (1 - \frac{1}{6})(1 - \frac{2}{11})(1 - \frac{4}{16})(1 - \frac{7}{11})(1 - \frac{1}{3}) = .1240$
Since $P(X > 500|X > 250) = \frac{S(500)}{S(250)}$ and $P(X > 6000|X > 250) = \frac{S(6000)}{S(250)}$
it follows that $P(X > 6000|X > 500) = \frac{S(6000)}{S(500)} = \frac{S(6000)/S(250)}{S(500)/S(250)}$.
Therefore, the estimate of $P(X > 6000|X > 500)$ is $\frac{.1240}{.8333} = .149$. Answer: B

27. The distortion risk measure is $\int_0^\infty g(S(t))\,dt$, where $S(t) = (\frac{\theta}{t+\theta})^\alpha = (\frac{1000}{t+1000})^4$ is the survival function of the Pareto. Since $g(x) = \sqrt{x} = x^{1/2}$, we have
$g(S(t)) = [S(t)]^{1/2} = (\frac{1000}{t+1000})^2$. This is the survival function for the Pareto with parameters $\theta = 1000$ and $\alpha = 2$. Then $\int_0^\infty g(S(t))\,dt = \int_0^\infty (\frac{1000}{t+1000})^2\,dt$. Since this is the integral of the survival function for the Pareto with $\theta = 1000$ and $\alpha = 2$, this integral is the mean of this Pareto. The mean of the Pareto with parameters $\theta = 1000$ and $\alpha = 2$ is $\frac{1000}{2-1} = 1000$.
Answer: D

28. From the graph, we can see that the data set consists of three data points,
The points are $x = 10$ with prob. .2, $x = 100$ with prob. .4, and $x = 1000$ with prob. .4.
The mean of the data is $(10)(.2) + (100)(.4) + (1000)(.4) = 442$.
The lognormal distribution has parameters μ and σ, and the distribution function of the lognormal is $F(x) = \Phi(\frac{ln\, x - \mu}{\sigma})$. We see from the graph that $F(10) = .2$ and $F(100) = .6$.
Therefore, $\Phi(\frac{ln\, 10 - \mu}{\sigma}) = .2$ and $\Phi(\frac{ln\, 100 - \mu}{\sigma}) = .6$.
From the normal table, with interpolation, we see that $\Phi(.253) = .6$. The normal table also gives us $\Phi(.842) = .8$, so that $\Phi(-.842) = .2$. This gives us the following two percentile equations: $\frac{ln\, 10 - \mu}{\sigma} = -.842$ and $\frac{ln\, 100 - \mu}{\sigma} = .253$. Dividing the second equation by the first results in $\frac{ln\, 100 - \mu}{ln\, 10 - \mu} = \frac{.253}{-.842} = -.3005$. Solving for μ results in $\mu = 4.07$, and then $\sigma = 2.10$. The mean of the lognormal is $e^{\mu + \frac{1}{2}\sigma^2} = e^{4.07 + \frac{1}{2}(2.1^2)} = 531$.
The difference between the two means is $531 - 442 = 89$.
Note that this is a matching percentile question. Answer: B

29. The quantity being estimated is the percentage of loss eliminated by applying a deductible of 2 to the insured loss. The total loss from the sample is 10 and a deductible of 2 would eliminate $\theta = 20\%$ of the insured loss. This is the empirical estimate based on the actual sample. Since fire and wind losses are independent and have different distributions, when we do resampling to find the Bootstrap estimate, we resample based on size 2 from fire losses, and we resample based on size 2 from wind losses. There are 4 possible resamplings of fire losses and 4 possible resamplings of wind losses.

Fire	Wind	Total Loss	Loss After Ded. of 2	Loss Eliminated, $\widehat{\theta}$	$(\theta - \widehat{\theta})^2$
$3,3$	$0,0$	6	6	0	$(.2-0)^2 = .04$
$3,3$	$0,3$	9	7	$\frac{2}{9}$	$(.2-\frac{2}{9})^2 = .0005$
$3,3$	$3,0$	9	7	$\frac{2}{9}$	$(.2-\frac{2}{9})^2 = .0005$
$3,3$	$3,3$	12	8	$\frac{4}{12}$	$(.2-\frac{4}{12})^2 = .0178$
$3,4$	$0,0$	7	7	0	$(.2-0)^2 = .04$
$3,4$	$0,3$	10	8	$\frac{2}{10}$	$(.2-.2)^2 = 0$
$3,4$	$3,0$	10	8	$\frac{2}{10}$	$(.2-.2)^2 = 0$
$3,4$	$3,3$	13	9	$\frac{4}{13}$	$(.2-\frac{4}{13})^2 = .0116$

$4,3$ is the same as 3,4 fire damage

$4,4$	$0,0$	8	8	0	$(.2-0)^2 = .04$
$4,4$	$0,3$	11	9	$\frac{2}{11}$	$(.2-\frac{2}{11})^2 = .0003$
$4,4$	$3,0$	11	9	$\frac{2}{11}$	$(.2-\frac{2}{11})^2 = .0003$
$4,4$	$3,3$	14	10	$\frac{4}{14}$	$(.2-\frac{4}{14})^2 = .0073$

The bootstrap estimate of the MSE of $\widehat{\theta}$ is $\frac{.04+.0005+\cdots+.0003+.0073}{16} = .013$. Answer: E

30. The Bayesian credibility combination of a prior β that is inverse gamma with parameters α and θ, and a model $X|\beta$ that is exponential with mean β, results in a posterior distribution that is also inverse gamma. If the number of observed values of x is n, and the values are $x_1, ..., x_n$ then the posterior inverse gamma distribution has parameters $\alpha' = \alpha + n$ and $\theta' = \theta + \Sigma x_i$. In this example, the prior is inverse gamma with $\alpha = 2$ and $\theta = c$. There is a single observed value of x, so $n = 1$. The posterior inverse gamma has parameters $\alpha' = \alpha + 1 = 3$ and $\theta' = \theta + \Sigma x_i = c + x$. The mean of the inverse gamma is $\frac{\theta}{\alpha - 1}$, so the mean of the posterior distribution is $\frac{c+x}{3-1}$.   Answer: C

31. For the Pareto distribution with known θ, the maximum likelihood estimate of α can be formulated as follows: $\widehat{\alpha} = \frac{n}{\Sigma ln(y_i+\theta) - \Sigma ln(d_i+\theta)}$,
where n is the number of uncensored observations, y_i is either the observed loss or the right censored loss (before deductible) for policy i and d_i is the deductible for policy i.
In this problem there are no censored losses, and all policies have a deductible of 100.
The mle of α is $\widehat{\alpha} = \frac{7}{ln(120+400)+\cdots+ln(2500+400)-7\,ln(100+400)} = 1.85$.
The expected loss of the Pareto is $\frac{\theta}{\alpha-1} = \frac{400}{.85} = 471$.   Answer: A

32. Condition A is not required. One of the features of Buhlmann-Straub is that there may be different numbers of exposures for each policy. Condition B is not required. The underlying distribution can have any distribution. Condition C is related to limited fluctuation credibility, not Buhlmann-Straub.   Answer: E

33. We can use this data as product limit data for which switching to a dieting program is considered a censored observation. The "death" points (reaching weight loss goal) are
$y_1 = 8$, $y_2 = 12$, $y_3 = 22$ and $y_4 = 36$.
The Nelson-Aalen estimate of $H(12) = H(y_2)$ is $\frac{s_1}{r_1} + \frac{s_2}{r_2}$.
There is a right-censoring at time 4, so at the first "death" point, $y_1 = 8$, there are $r_1 = 7$ at risk and $s_1 = 1$ "deaths" (there is a right-censoring at time, but since it is a tie with the death time, it is held over until the next death point). At death time $y_2 = 12$, there are $r_2 = 5$ at risk (the original 8 minus the death and the two censorings) and there are $s_2 = 2$ deaths (again there is a censoring at time 12, but since this is tied with the death time, the censoring is not counted until the next death point). The Nelson-Aalen estimate is $\widehat{H}(12) = \frac{1}{7} + \frac{2}{5} = .5429$. The estimated variance of the Nelson-Aalen estimate is $V\widehat{a}r[\widehat{H}(12)] = \frac{s_1}{r_1^2} + \frac{s_2}{r_2^2} = .1004$.
The upper limit of the 90% linear confidence interval for $\widehat{H}(12)$ is
$.5429 + 1.645\sqrt{.1004} = 1.06$.   Answer: D

34. According to the procedure described in the McDonald text, the annualized expected return on the stock is $ln(\frac{S_1}{S_0}) + .5\sigma^2$, where $ln(\frac{S_1}{S_0})$ is the natural log of the growth in the stock price for one year and σ^2 is the variance of $ln(\frac{S_1}{S_0})$. We follow the procedure in Example 18.8 on page 606 of the text. From the given table, we can find $ln(\frac{S_t}{S_{t-1}})$, where t is measure in months:

t	$ln(\frac{S_t}{S_{t-1}})$
1	$ln(\frac{56}{54}) = .036368$
2	$-\ .154151$
3	$.136132$
4	$.087011$
5	$-\ .033902$
6	$.066691$

We can estimate the monthly mean of the log price ratio as the sample mean of these values. This is $.023025$. The estimated annual mean of the log price ratio is $.023025 \times 12 = .2763$.
We can estimate the variance of the monthly log price ratio using the unbiased sample variance estimate of these 6 values. This is $.01072$. The estimated variance of the annual log price ratio is $.01072 \times 12 = .12864$. The estimated annualized expected return is
$.2763 + .5(.12864) = .341$. Answer: E

35. $P(G = \frac{1}{3}|D = 0) = \frac{P(G=\frac{1}{3}\cap D=0)}{P(D=0)}$.
$P(G = \frac{1}{3} \cap D = 0) = P(D = 0|G = \frac{1}{3}) \cdot P(G = \frac{1}{3}) = \frac{1}{3} \cdot \frac{2}{5} = \frac{2}{15}$.

$P(D = 0) = P(D = 0 \cap G = \frac{1}{3})+P(D = 0 \cap G = \frac{1}{5})$
$= P(D = 0|G = \frac{1}{3}) \cdot P(G = \frac{1}{3}) + P(D = 0|G = \frac{1}{5}) \cdot P(G = \frac{1}{5})$
$= \frac{1}{3} \cdot \frac{2}{5} + \frac{1}{5} \cdot \frac{3}{5} = \frac{19}{75}$.

$P(G = \frac{1}{3}|D = 0) = \frac{P(G=\frac{1}{3}\cap D=0)}{P(D=0)} = \frac{2/15}{19/75} = \frac{10}{19}$. Answer: E

36. We apply the usual Buhlmann method to the $n = 80$ $(24 + 30 + 26 = 80)$ exposures.
The hypothetical means are given in the first table. The expected hypothetical mean is $(2000)(.6) + (3000)(.3) + (4000)(.1) = 2500$. This is μ.
The process variance is 1000^2 for each class, so the expected process variance is $v = 1000^2$.
The variance of the hypothetical mean is the variance of the 3-point variable for θ. This variance is $a = [(2000)^2(.6) + (3000^2)(.3) + (4000^2)(.1)] - 2500^2 = 450,000$.
The Buhlmann factor Z is $Z = \frac{80}{80+\frac{1000^2}{450,000}} = .9730$.
The sample mean of the 80 exposures is $\frac{24,000+36,000+28,000}{24+30+26} = 1100$.
The Buhlmann credibility estimate for the loss per exposure in Year 4 is
$Z\overline{X} + (1 - z)\mu = (.973)(1100) + (.027)(2500) = 1137.8$. Answer: B

37. We are trying to estimate a probability q using a sample of size n, with estimator
$\widehat{q} = \frac{\text{\# sample values} \leq 300}{n}$. We know the actual value of q is
$q = P(X \leq 300) = 1 - e^{-300/100} = .950213$.
The number of values needed is $n \geq (\frac{2.576}{.01})^2 \cdot \frac{1-q}{q} = (\frac{2.576}{.01})^2 \cdot \frac{1-.950213}{.950213} = 3477$.
The value of 2.576 is the 99.5-th percentile of the standard normal distribution.
If we didn't know the exact value of q, we would use $n \geq (\frac{2.576}{.01})^2 \cdot \frac{n-P_n}{P_n}$,
where $\widehat{q} = \frac{P_n}{n}$. Answer: E

38. $S_n(y_4) = S_n(y_3)(1 - \frac{s_4}{r_4})$. Therefore, $.50 = (.65)(1 - \frac{3}{r_4})$.
It follows that $r_4 = 13$. Since there are 3 deaths at time y_4 and 6 censorings between y_4 and y_5, the number at risk at time 5 must be $r_5 = 13 - 3 - 6 = 4$.
Then, from $S_n(y_5) = S_n(y_4)(1 - \frac{s_5}{r_5})$ we get $.25 = (.5)(1 - \frac{s_5}{4})$, it follows that $s_5 = 2$.
Answer: B

39. The expected number of losses from the negative binomial frequency distribution is $r\beta = 15$. The probability that a loss is above the deductible is
$P(X > 200) = e^{-(200/\theta)^\tau} = e^{-(200/1000)^{.3}} = .5395$.
Out of the 15 losses that occur on average, the expected number that are above the deductible of 200 (and result in an insurance payment being made) is $15(.5395) = 8.1$. Answer: C

40. q is the probability that a single policy has 1 or more claims. The sample estimate of q is
$\widehat{q} = \frac{400}{2000} = .2$. The variance of $\widehat{q}$ is $\frac{q(1-q)}{n}$, and the estimated variance of $\widehat{q}$ is
$\frac{(.2)(.8)}{2000} = .00008$. The upper bound of the 95% confidence for q is
$\widehat{q} + 1.96\sqrt{Var(\widehat{q})} = .2 + 1.96\sqrt{.00008} = .21753$. Answer: D