losses (including the deductible)

were 1000, 2000, 4000, 8000 and two losses above 10000. The loss distribution is fitted to an exponential. Find the MLE of θ.

$d_i = d = 500$

Soln: $\hat{\theta} = \dfrac{\sum_{i=1}^{n}(x_i - d_i)}{m} =$

$$\frac{500 + 1500 + 3500 + 7500 + 2(9500)}{4} = \boxed{8000}$$

see HWS 1. Hw4 3b

75] If $L(\theta) = \theta^{-a} e^{-\frac{b}{\theta}}$, $\hat{\theta} = \frac{b}{a}$.

In 7(a), $a = m$ and $b = \sum_{i=1}^{n}(x_i - d_i)$.

ex] $L(\theta) = \dfrac{e^{-4250/\theta}}{\theta^{17}}$, $\hat{\theta} = \dfrac{4250}{17} = 250$.

76] If $L(\lambda) = \lambda^a e^{-\lambda b}$, $\hat{\lambda} = \frac{a}{b}$.

(In 75), $\lambda = \frac{1}{\theta}$ so $\hat{\lambda} = \frac{1}{\hat{\theta}}$.)

ex] $X \sim U(0,\theta)$. 10 observations are below 90 and 1 above (or equal to 90 due to censoring).

$\hat{\theta} = \frac{n}{m} c_j = \frac{11}{10} 90 = \boxed{99}$. see HWS 5.

ex]

range	number of obs's
(0, 1000)	18
(1000, 5000)	32
(5000, 10000)	30
(10000, ∞)	0

$\left.\begin{matrix}18\\32\end{matrix}\right\} m = 50$

$P(X = c) = 0$ if $X \sim U(0, \theta)$ is uncensored

(from 74) $c = 5000$, $n = 80$, $m = 50 = \#$ obs's $< c$

$$\hat{\theta} = \min\left(10000,\ 5000\ \frac{80}{50}\right) = 8000$$

77] The Bernoulli technique is useful if there are 2 classes or groups and $p = p(\theta) = P(X \text{ in 1st group})$ has $\theta = p^{-1}(p(\theta))$.

0	(0, c)	n_1	p
1 or more	(c, ∞)	n_2	$1-p$
	sum	n	1

Then $L(p) = p^{n_1}(1-p)^{n_2}$, $\hat{p} = \frac{n_1}{n}$, $1-\hat{p} = \frac{n_2}{n}$.

Solve $p \stackrel{\text{set}}{=} \hat{p} = \frac{n_1}{n}$ for θ. (works by invariance).

For class $1 = \{0\}$, $p = p_0$. For class $(0, c)$, $p = F(c)$.

78] If $L(\tau) = \tau^{n_1}(1-\tau)^{n_2}$, then the MLE $\hat{\tau} = \frac{n_1}{n_1 + n_2}$. (Bernoulli likelihood)

X is 0,1 data

ex] claim counts follow a Poisson dist

number of claims	number of policyholders
0 claims	65
1 or more claims	35

use $p_0 \stackrel{\text{set}}{=} \hat{p}_0$

a) Find the MLE of λ.

b) Find the MLE of $P_2 = P(X=2)$.

Soln a) $P_0 = e^{-\lambda} \stackrel{set}{=} .65 = \hat{P}$ see HW5 #2

so $-\hat{\lambda} = \ln(.65)$, $\underbrace{\hat{\lambda} = -\ln(0.65)}$ $= \boxed{0.4308}$

b) $P_2 = \frac{e^{-\lambda}\lambda^2}{2!} = e^{-\lambda}\frac{\lambda^2}{2}$, $\hat{P}_2 = 0.65 \frac{[\ln(.65)]^2}{2} =$

$\boxed{0.06031}$ by invariance.

79] If there are K parameters and K+1 classes, the MLES $\hat{\theta}_1, \ldots, \hat{\theta}_K$ are such that $P_i = Prob(X \text{ in class } i)$ with $\hat{P}_i = \frac{n_i}{n}$, $i = 1, \ldots, K+1$

one p_i is redundant

(c_0, c_1)	n_1	P_1
(c_1, c_2)	n_2	P_2
$\vdots$	$\vdots$	
(c_K, c_{K+1})	n_{K+1}	P_{K+1}
sums	n	1

$n = n_1 + \cdots + n_{K+1}$

Solve $P_i \stackrel{set}{=} \frac{n_i}{n}$, $i = 1, \ldots, K$ for $\hat{\theta}_1, \ldots, \hat{\theta}_K$.

Here $P_i = F(c_i) - F(c_{i-1})$.

75) is the special case with $K=1$.

ex}

range	# claims
(0 – 1000)	65
(1000 – 2000)	30
over 2000	5

} sum to n = 100

35.5

Then $p_1 = F(1000) \overset{set}{=} 0.65 = \hat{p}_1$

$p_2 = F(2000) - F(1000) \overset{set}{=} .3 = \hat{p}_2$

and $S(2000) = 1 - \hat{p}_1 - \hat{p}_2 = .05$

ex} 535 claims 1000 or less and 465 claims 1000 or more. Fit an EXP(θ) dist and estimate $\hat{\theta}$.

soln!

(0 – 1000)	535
(1000, ∞)	465

} sum to 1000

use k = 1

$$p = F(1000) = 1 - e^{\frac{-1000}{\theta}} \overset{set}{=} 0.535 = \hat{p}$$

$$e^{\frac{-1000}{\theta}} = .465$$

$$\frac{-1000}{\theta} = \ln(0.465)$$

$$\hat{\theta} = \frac{-1000}{\ln(0.465)} = \boxed{1305.9640}$$

could have used $S(1000) \overset{set}{=} .465$

Skip § 15.5 Bayesian methods for now. We will do §15.5 before credibility.

§ 15.6 Estimation for Discrete Distributions

80] Did MME, MLE, $I_n(\theta)$, $var(\hat{\theta})$, $var(g(\hat{\theta}))$, CI for $\hat{\theta}$, etc.

81] The Poisson, binomial, negative binomial and Geometric distributions are the only members of the $(a,b,0)$ class. X is a member of the $(a,b,0)$ class if

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k} \quad \text{for } k = 1, 2, \ldots, \quad \text{except}$$

the recursion goes up to $k = m$ for the binomial.

Hence $\dfrac{k\,p_k}{p_{k-1}} = ak + b$ for $k = 1, 2, \ldots$.
(not for $k = 0$)

In a sample (eg 0,1,1,5,0,3,7,0,5,2,1,1,1,4,2,2), let n_k = number in sample equal to k with $n = \sum_k n_k$. Plot k vs $\dfrac{k\,\hat{p}_k}{\hat{p}_{k-1}} = \dfrac{k\,n_k}{n_{k-1}}$ where k is omitted if $n_k = 0$. If the n_k are large, the plot should follow a straight line with slope a.

The Poisson RV X has slope $a = 0$ and $E(X) = V(X)$.

The binomial RV X has slope $a < 0$ and $E(X) > V(X)$. $V(X) < E(X)$

The negative binomial RV X (geometric $= NB(\beta, r=1)$) has $a > 0$ and $E(X) < V(X)$. $V(X) > E(X)$ Often use N instead of X.

dist	a	b	p_0	
Poisson(λ)	0	λ	$e^{-\lambda}$	$E(X) = V(X)$
bin(q, m)	$\frac{-q}{1-q}$	$(m+1)\frac{q}{1-q}$	$(1-q)^m$	$E(X) > V(X)$
$NB(\beta, r)$	$\frac{\beta}{1+\beta}$	$(r-1)\frac{\beta}{1+\beta}$	$(1+\beta)^{-r}$	$E(X) < V(X)$
Geom(β)	$\frac{\beta}{1+\beta}$	0	$(1+\beta)^{-1}$	$E(X) < V(X)$

Check with $\bar{X}$, $\hat{\sigma}_E^2$ or $\hat{\sigma}_U^2$

$$\frac{k p_k}{p_{k-1}} = \frac{k e^{-\lambda} \lambda^k / k!}{e^{-\lambda} \lambda^{k-1}/(k-1)!} = \lambda \text{ for poisson}$$

ex}

# claims	# policies
0	672
1	660
2	260
3	55
4	7
5	1
6+	0
total	1655

which distribution from the (a, b, 0) class best fits the data?

$$\frac{n_1}{n_0} = \frac{660}{672} = .9821$$

$$\frac{2n_2}{n_1} = \frac{2(260)}{660} = .7879$$

$$\frac{3n_3}{n_2} = \frac{3(55)}{260} = .6346$$

$$\frac{4n_4}{n_3} = \frac{4(7)}{55} = .5091$$

slope is negative so binomial.

over for plot

n_4 and n_5 are too small $\rightarrow \frac{5n_5}{n_4} = .7143$

or

$$\bar{X} = \frac{660 + 260(2) + 55(3) + 7(4) + 1(5)}{1655} = 0.8326$$

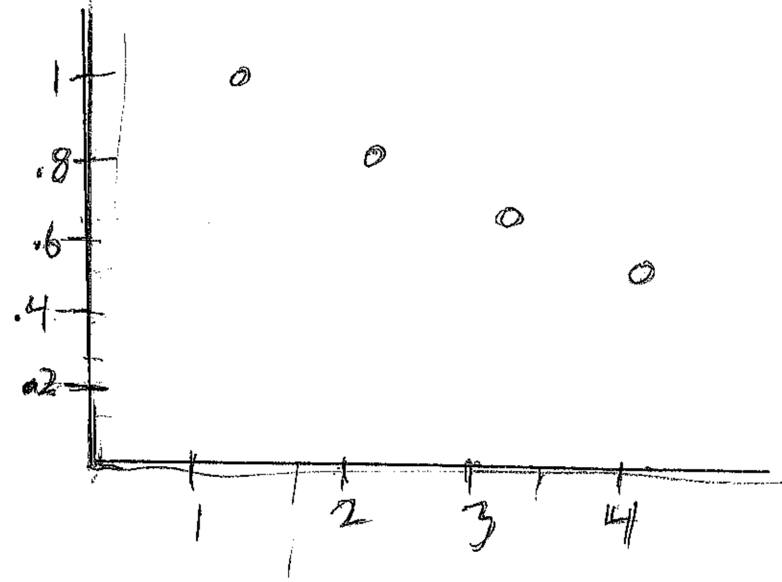
$$= \frac{\Sigma k n_k}{\Sigma n_k} = m = \bar{x}$$

$$t = \frac{\Sigma x_i^2}{n} = \frac{660 + 260(2^2) + 55(3^2) + 7(4^2) + 1(5^2)}{1655} = 1.4091$$

$$= \frac{\Sigma k^2 n_k}{n}$$

$$\hat{\sigma}_E^2 = t - m^2 = 1.4091 - (.8326)^2 = 0.7158$$

$$\hat{\sigma}_U^2 = \frac{n}{n-1}\hat{\sigma}_E^2 = \frac{1655}{1654} \cdot .7158 = 0.7163 < \bar{X}$$ so binomial



Should be close between

32) For Poisson, won't get $\bar{X} = \hat{\sigma}_E^2$ or $\bar{X} = \hat{\sigma}_U^2$,

but $\dfrac{k\,\hat{p}_k}{\hat{p}_{k-1}} \simeq$ constant if both n_k and n_{k-1} are large (oscillate about a constant horiz line rather than clearly increase or decrease).

33)

a)

k	n_k
0	n_0
1	n_1
2	n_2
$\vdots$	$\vdots$
m	n_m
$(m+1)+$	0

$$L = \prod_{k=0}^{m} p_k^{n_k}$$

$$\ln L = \sum_{k=0}^{m} n_k \ln(p_k)$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i^j = \frac{1}{n}\sum_{k=0}^{m} k^j n_k. \quad \text{Usually } j = 1, 2.$$

b)

k	n_k
0	n_0
$\vdots$	$\vdots$
m	n_m
$(m+1)+$	$n_{m+1}^+ > 0$

$$L = \left(\prod_{k=0}^{m} p_k^{n_k}\right)\underbrace{\left(1 - p_0 - \cdots - p_m\right)}_{P(X \geq m+1) = 1 - P(X < m+1)}^{n_{m+1}^+}$$

$$= 1 - P(X \leq m)$$

$$p_k = P(X = k)$$

Skip the (a, b, 1) class. Skip 12.6, 12.7.

Ch 16

p442-3

1) Let X have cdf F, pdf f.

If X is left truncated at d, let

$$F^*(x) = \frac{F(x) - F(d)}{1 - F(d)} \quad \text{and} \quad f^*(x) = \frac{f(x)}{1 - F(d)}$$

be the cdf and pdf of the truncated RV for $x \geq d$.

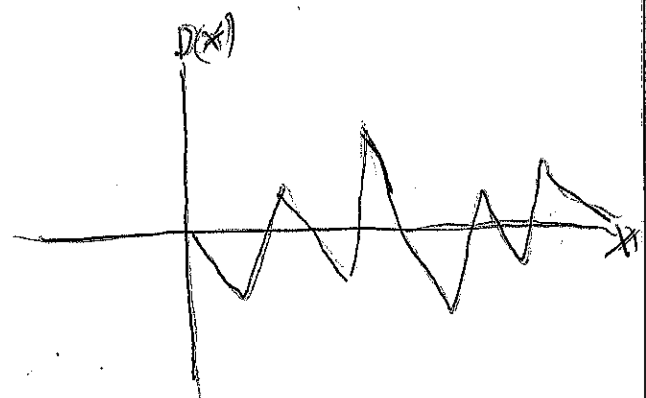
Let the empirical cdf $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x)$

where $I(X_i \leq x) = \begin{cases} 1 & X_i \leq x \\ 0 & X_i > x, \end{cases}$

and let the empirical pdf be $f_n(x)$.

2] Suppose $X_1, \ldots, X_n$ are iid (individual data). A $D(x)$ plot is a plot of x vs $D(x) = F_n(x) - F^*(x)$

where $F^*(x) = \hat{F}^*(x)$ is often found using MLEs or MMEs. Want the plot to oscillate about the $D(x) = 0$ line.



Note that $F_n(d) - F^*(d) = 0$

where $d = 0$ if there is no truncation.

3] * Know The p-p plot is used for $X_1, \ldots, X_n$ iid.

Order the observations

$$\underbrace{x_1 \le x_2 \le \cdots \le x_n}_{\text{actually order statistics}} .$$

Plot "$F_n(x_j) = \frac{j}{n+1}$" vs $\underbrace{F^*(x_j)}_{\text{often from MLE}}$

(Using 1) gives $F_n(x_j) = \frac{j}{n}$, hence the quotes.)

The x and y axes go from 0 to 1.

Horizontally use multiples of $\frac{1}{n+1}$ unless there is a tie.

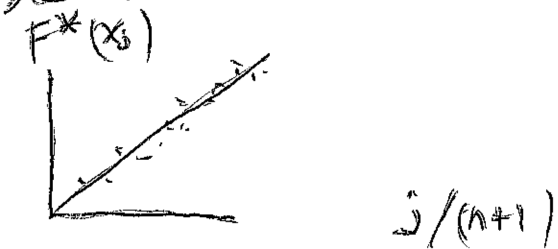
If there is a tie among the x_j's $\underbrace{x_i, \dots, x_{j_{max}}}_{k \text{ tied}}$

could use $F_n(x_j) = \frac{j_{max}}{n+1}$ for $j = i, \dots, j_{max}$

or use the average $\frac{\overbrace{x_i + \cdots + x_{j_{max}}}^{\text{just } x_i \text{ so same}}}{k}$ vs $\frac{j_{max}}{n+1}$.

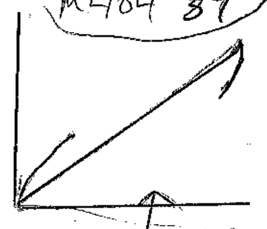
Censored values do not get plotted.

4) Good fit means plotted points are close to the identity line through (0,0), (1,1).



M404 39

5)
a) Slope of curve in PP plot > 1 in region (especially (tail))



$\Rightarrow$ thick or heavy or fitted distribution has more probability in that region than the sample or too much weight in region

heavy tails since slope >1 in the 2 tail regions

b) Slope of curve in PP plot < 1 in region (tail)

$\Rightarrow$ thin or light or fitted distribution (F^*) has less probability in that region than the sample (F_n) or too little weight in region

c) left tail: ~~region where~~ , right tail region where

$\frac{j}{p+1} \approx 0$, $\frac{j}{p+1} \approx 1$

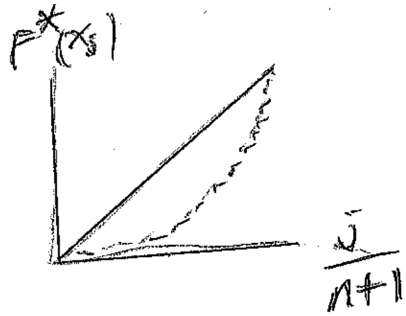
(overlap on plots)

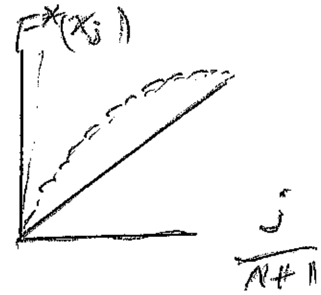
6) When testing H_0 vs H_1 or H_A we do not want to reject the null hypothesis H_0 unless it is very unlikely that we would observe what we did from the sample if the null hypothesis is true.

7) H_0 fitted distribution is good H_A fitted distribution is not good

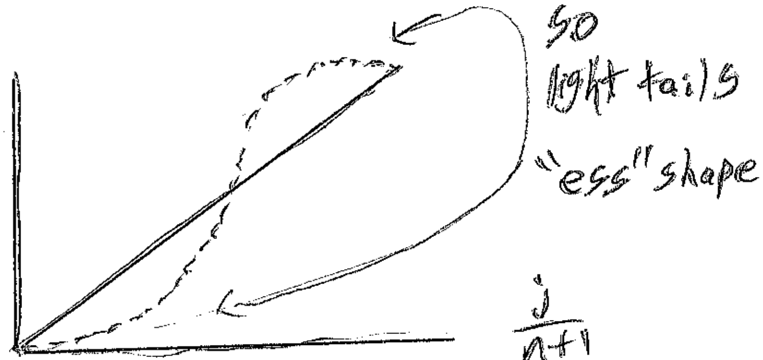
8) χ^2 test statistic $Q = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}$

8) b continued on next page





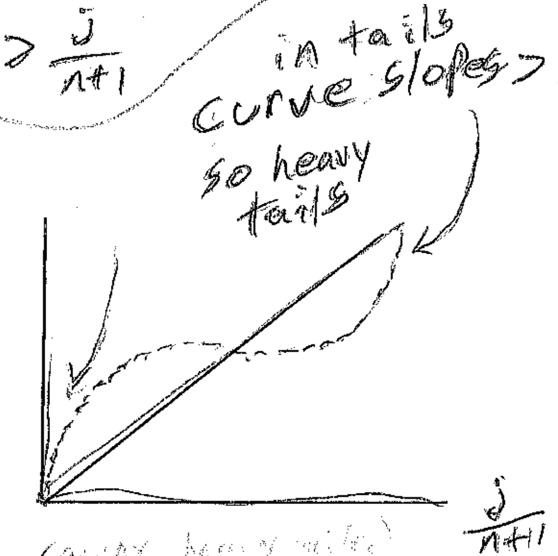
$F^*(x_j) < \frac{j}{n+1}$ $\qquad$ $F^*(x_j) > \frac{j}{n+1}$



normal, uniform light-tailed

$F^*(x_j) < \frac{j}{n+1}$ then $F^*(x_j) > \frac{j}{n+1}$

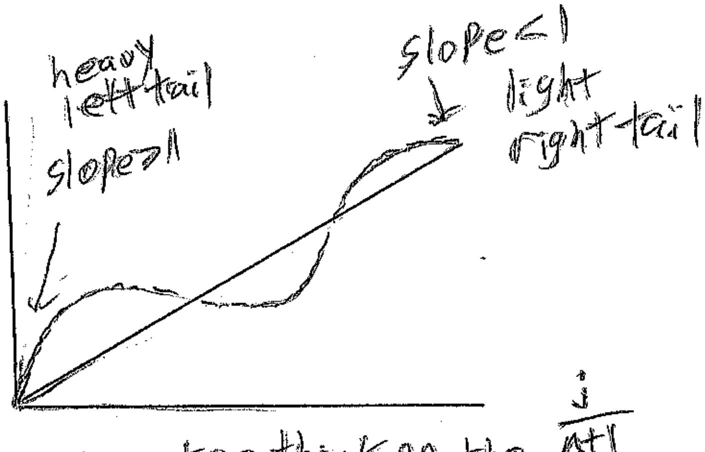
left and right tails of fitted distribution are too thin or light



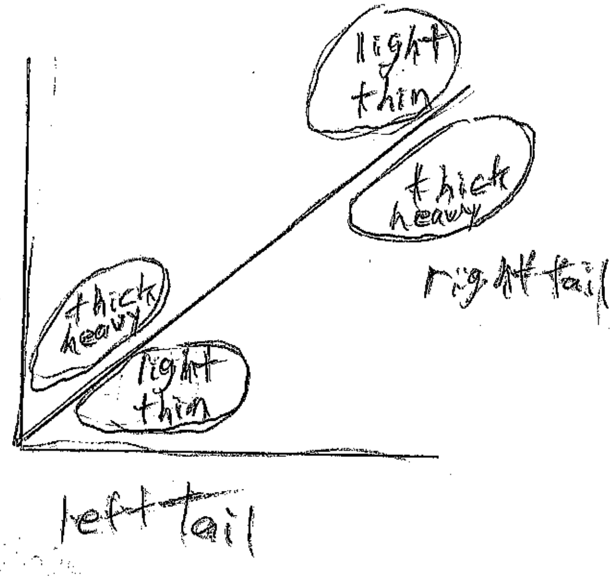
cauchy heavy tailed

$F^*(x_j) > \frac{j}{n+1}$ then $F^*(x_j) < \frac{j}{n+1}$

left and right tails of fitted distribution are too thick or heavy



tail is too thick on the left, too thin on the right
fitted distribution has less probability than the sample near the middle (or median)



group	prob $p_i = F(c_i) - F(c_{i-1})$, F fitted					or		
(c_0, c_1)	p_1	n_1	0	p_1	n_1	0	p_1	$\leftarrow$ group 1
(c_1, c_2)	p_2	n_2	1	p_2	n_2	1	p_2	
$\vdots$			2	p_3	n_3	2	p_3	
		or	$\vdots$		$\vdots$	$\vdots$		
						$k-2$	p_{k-1}	
(c_{k-1}, c_k)	p_k	n_k	$k-1$	$\underbrace{p_k}$	n_k	$(k-1)+$	$(1 - p_1 - \cdots - p_{k-2}) = p_k$	

$\uparrow$ $c_k = \infty$ possible

not using $p_i = P(X=i)$.

Then $O_i = n_i$ and $E_i = n p_i$ where $n = \sum_{j=1}^{k} n_i$.

9] Using χ^2 table: let

degrees of freedom $= d = k - r - 1$

where $r =$ number of parameters estimated (preferably by MLE), and $r = 0$ is possible.

Skip

10) If $d > 20$! if $W \sim \chi^2_d$ then

$$\sqrt{2W} - \sqrt{2d-1} \approx N(0,1) \quad \text{so}$$

$$\alpha \approx P(\sqrt{2X} - \sqrt{2d-1} \le z_\alpha) = P(\sqrt{2X} < z_\alpha + \sqrt{2d-1})$$

$$= P[2W \le (z_\alpha + \sqrt{2d-1})^2] = P[W \le 0.5(z_\alpha + \sqrt{2d-1})^2]$$

So $\chi^2_d(\alpha) \approx 0.5\,(z_\alpha + \sqrt{2d-1})^2$.

Note $0.5\,(1.645 + \sqrt{2(20)-1})^2 = 31.1260 \approx \underbrace{31.41}_{\text{table}}$

10) <u>Know</u> The 4 step χ^2 test of hypotheses:

i) Ho the fitted distribution is good H_A not Ho

ii) $$Q = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}$$

iii) df = k - r - 1

If Q > cutoff then pval < α
reject Ho. If Q < cutoff then pval > α
fail to reject Ho.

iv) "nontechnical" conclusion
reject Ho → fitted distribution is <u>not</u> good

fail to reject Ho → fitted distribution is good
(or not enough evidence to say fitted dist
is not good).

11) * Use $\alpha = 0.05$ if α is not given.

α = significance level

12] χ^2 table

df	.9 ($\alpha = .1$)	.95 ($\alpha = .05$)	.975 ($\alpha = .025$)
1	2.706	3.841	5.024
2	4.605	5.991	7.378
3	6.251	7.815	9.348
4	7.779	9.488	11.143
⋮			

critical values = cutoff

If df = 3 and Q = 7 reject at 0.1 significance level but not at 0.05 significance level.

reject H_0 if $Q >$ critical value so pval $< \alpha$

13]

category (given)	$O_i = n_i$ (given)	p_i (sometimes given)	$E_i = np_i$	χ^2 contrib $= \frac{(O_i - E_i)^2}{E_i}$
1	n_1	p_1	E_1	c_1
2	n_2	p_2	E_2	c_2
⋮	⋮	⋮	⋮	⋮
k	n_k	p_k	E_k	c_k
	$n = \sum_{i=1}^{k} n_i$	$1 = \sum_{i=1}^{k} p_i$ up to rounding	$n = \sum E_i$	$Q = \sum_{i=1}^{k} c_i$

p_i = fitted prob X is in ith category. could be given, if category $1 = \{1\}$, $p_1 = \hat{p}_1$

if category $1 = (c_0, c_1)$, $p_1 = \hat{F}(c_1) - \hat{F}(c_0)$.

plug in MLE $\hat{\theta}$ if θ is not given ex C71] Investigating insurance fraud. (41.5

Prob's for when there is no fraud are given.

Test Ho no fraud.

$E_i = np_i$ χ^2 contrib $= \dfrac{(O_i - E_i)^2}{E_i}$

# claimants per accident	standard prob	obs # accidents	E_i	χ^2 contrib
1	0.25	235	250	$\dfrac{(235-250)^2}{250} = 0.9000$
2	0.35	335	350	0.6429
3	0.24	250	240	0.4167
4	0.11	111	110	0.0091
5	0.04	47	40	1.225
6+	0.01	22	10	14.4000
	1	n=1000		Q = 17.5937

given

i) Ho fitted distribution is good HA not Ho

ii) Q = 17.5937

iii) $df = k - r - 1 = 6 - 0 - 1 = 5$

	.05	.005
	.95	.995
5	11.070	16.750

Q > 11.070 reject Ho (α = .05 if not given)

iv) there is fraud

ex] know variant Q = 16.25 df = 5

	.01	.005
	.99	.995
5	15.086	16.75

reject at the 0.01 significance level

but not at the 0.005 significance level