Abraham Nieto 51556

1 a)
si hacemos las suma de las probabilidades marginales podemos ver que tanto para padres(renglones) como para hijos(columnas) no son iguales por tanto no son homogéneas

	a	b	c	d	e	suma
a	0.018	0.035	0.031	0.008	0.018	0.11
b	0.002	0.112	0.064	0.032	0.069	0.279
c	0.001	0.066	0.094	0.032	0.084	0.277
d	0.001	0.018	0.019	0.01	0.051	0.099
e	0.001	0.029	0.032	0.043	0.13	0.235
suma	0.023	0.26	0.24	0.125	0.352	

1 b)
Dado que lo que queremos modelar son las probabilidades de que los padres o hijos tengan cierta ocupación y son 5 categorías los mas apropiado sería modelar con una distribución Multinomial de tal forma que la función de probabilidad esta dada por: $\frac{n!}{x_1!x_2!x_3!\ldots x_k} p_1^{x_1} p_2^{x_2} p_3^{x_3} \ldots p_k^{x_k}$

donde k=5 y $x_i = ocupacion_i$ y p_i es la suma d las marginales de la tabla que se muesta en el inciso a) y con esto la varianza sería $np_i(1-p_i)$ con $x_i=1$ si está en la categoría i y 0 en otro caso.

De modo que $P(X_1=x_1,\ldots,X_5=x_5)=\frac{n!}{x_1!\ldots x_5!}(.023)^{x_1}(.26)^{x_2}(.24)^{x_3}(.125)^{x_4}(.352)^{x_5}$ para los hijos y

para los padres $P(X_1=x_1,\ldots,X_5=x_5)=\frac{n!}{x_1!\ldots x_5!}(.11)^{x_1}(.279)^{x_2}(.277)^{x_3}(.099)^{x_4}(.235)^{x_5}$

entonces las varianzas por ocupación serían:
hijos:

$var(x_i=a)=1e6*(.023)(.977)=22471$
$var(x_i=b)=1e6*(.26)(.74)=192400$
$var(x_i=c)=1e6*(.024)(.76)=182400$
$var(x_i=d)=1e6*(.125)(.875)=109375$
$var(x_i=e)=1e6*(.352)(.648)=228096$

Y $\sum_{i=0}^{5} p_i=1$ y suponemos que las marginales son homogéneas para aplicarlo así.

1 c)
En este caso lo que queremos ver es la probabilidad condicional $P(x_i=i : y_i=i)$ donde y_i son las ocupaciones de los padres y x_i la de los hijos entonces podemos entonces el modelo que puede hacer aquí es pensar en 2 categorias {iguales, distintos} hablando de ocupaciones en este caso podemos ver que la la proba de que sean las mismas ocupaciones es $\sum_{i=1}^{5} P_{ii}=.364$ que es la proba de que padres e hijos tengan la misma ocupación entonces

$P(x_i = i : y_i = j) = P(x_i = i, y_i = i)(1 - P(x_i = i, y_i = i)) = (\sum_{i=1}^{5} p_{ii})^{x_i}(1 - \sum_{i=1}^{5} p_{ii})^{1-x_i}$ con $X_i \in igual, distinto$

podemos ver que e modelo a usar aquí es una Bernoulli con parámetro $\sum_{i=1}^{5} P_{ii} = .364$

entonces la pdf se ve $P(x_i = i : y_i = j) = (.364)^{x_i}(.636)^{1-xi}$

1 d)

En este caso tenemos información suplementaria de tal modo que que lo que estamos buscando es $P(x_i = i : suplemento) = Dir(p_1, p_2, \dots p_5 : a_1, a2, \dots a_5) aprox \prod_{i=1}^{5} p_i^{(a_j - 1)}$

donde $a_1 = a_2 = a_3 = a_4 = .3\, y\, a_5 = .6$

1 e)

Dado que ahora queremos combinar la información de los datos y del suplemento entonces para predecir un nuevo caso usaremos una dirchlet multinomial de tal forma que $P(x_i = i : datos, suplemento) = \sum Mult(x_i : p_i) * Dir(p_i : a_i)$ donde P_i son los valosres de la tabla y a_i son los valores suplementarios definidos en el ejercicio anterior.

Dado que queremos predecir que los hijos tengan ocupación distinta al padre tenemos entonces Z={distintos,iguales} las categorias entonces podemos modelar lqa categoría i como: $P(Z_i = i) = q_i^{z_i}(1 - q_i)^{1-z_i}$ donde $q_i = \sum_{i \neq j} p_{ij}$ entonces combinando con el inciso anterior tenemos las siguientes probabilidades:

$P(Z_i = distinto : padre = a) = (.6)(.982)(.018)^0$

$P(Z_i = distinto : padre = b) = (.6)(.888)(.112)^0$

$P(Z_i = distinto : padre = c) = (.6)(.906)(.094)^0$

$P(Z_i = distinto : padre = d) = (.6)(.99)(.01)^0$

$P(Z_i = distinto : padre = e) = (.3)(.87)(.013)^0$

1 f)

Para caso donde se requiere la propensión de la ocupación de los padres podemos plantear el modelo de la misma forma que lo hicimos en el inciso c) lo único que cambia es el orden de las variables entonces tenemos que $P(y_i = i : x_i = i)$ en este caso debemos enfatizar que las $p_i ' s$ son ahora las p_{ji} 's de la matriz de probas ya que queremos la probas sobre los renglones.

1 g)

para la ocupación de los padres tenemos lo siguiente:

$p(y_i = i : x_i = a)$ para $i = a, b, c, d, e$ entonces tenemos que $p(y_i = i : x_i = a) aprox \prod_{i=1}^{5} p_{ij}^{(y_i)}$ donde j=a entonces tenemos que por ejemplo la proba de que el padre sea artesano dado que el hijo lo es :

$$p(y_a = 1, y_b = 0, y_c = 0, y_d = 0, y_e = 0 : x_i = a) = \frac{(.018)^1(.002)^0(.001)^0(.001)^0(.001)^0}{0.023} = 0.782$$

de que sea b=granjero

$$p(y_a=0, y_b=1, y_c=0, y_d=0, y_e=0 : x_i=a)=\frac{(.018)^0(.002)^1(.001)^0(.001)^0(.001)^0}{0.023}=0.086$$

de que sea c=obrero

$$p(y_a=0, y_b=0, y_c=1, y_d=0, y_e=0 : x_i=a)=\frac{(.018)^0(.002)^0(.001)^1(.001)^0(.001)^0}{0.023}=0.043$$

de que sea d=vendedor

$$p(y_a=0, y_b=0, y_c=0, y_d=1, y_e=0 : x_i=a)=\frac{(.018)^0(.002)^0(.001)^0(.001)^1(.001)^0}{0.023}=0.043$$

de que sea e=tecnico

$$p(y_a=0, y_b=0, y_c=0, y_d=0, y_e=1 : x_i=a)=\frac{(.018)^0(.002)^0(.001)^0(.001)^0(.001)^1}{0.023}=0.043$$

2 a)

De acuerdo con el resultado de la sección 7.6 tenemos que l es el % de fortuna entonces $\frac{l_i f}{x_i}$ es el gasto en A_i entonces $U[x]=\sum q_i \log(l_i f)-\sum q_i \log(x_i)=\sum q_i[\log(l_i)+\log(f)-\log(x_i)]$ entonces la utilidad esperada es de la forma $\log(f)+\sum q_i \log(l_i)-\sum q_i \log(x_i)$

2 b)

Tenemos que $P\{S=s|A_i\}=p_{s,i}$. pero también sabemos que $P\{A_i|S\}=\frac{P(S|A_i)P(A_i)}{\sum_{k=1}^{n} P(S|A_k)P(A_k)}=\frac{p_{s,i}q_i}{\sum p_{s,k}q_k}$ entonces si definimos $q_i^{'}=\frac{p_{s,i}q_i}{\sum p_{s,k}q_k}$ que obviamente esta en [0,1] entonces podemos probar claramente que $\sum q_i^{'}=\sum_i \frac{p_{s,i}q_i}{\sum p_{s,k}q_k}=1$

2 c)

Sabemos por el inciso a) que $E[\sum q_i \log(q_i)]-E[\sum q_i \log(x_i)]$ entonces sabemos que la proba aumenta a $q_i^{'}$ y tenemos esta aumento representado por $d(q_i, q_i^{'})$ entonces la utilidad esperada cambia a $E[\sum(q_i+d(q_i,q_i^{'}))\log(q_i+d(q_i,q_i^{'}))]-E[\sum(q_i+d(q_i,q_i^{'}))\log(x_i)]$ que de acuerdo con lo visto en el inciso a) tenemos que $E[U[X]]=\iiint U(X)p(p_s, q, x)d p_s dq d\$x$

3.

El paper habla sobre la diferencia entre las estadísticas bayesianas y frecuentistas al hacer afirmaciones sobre la relación entre los valores observables. Mostramos cómo los modelos estándar bajo ambos paradigmas se puede basar en un supuesto de intercambiabilidad y derivamos covarianza y correlación útiles resultados para valores de una secuencia intercambiable. Encontramos que tales valores nunca están correlacionados negativamente, y generalmente están correlacionados positivamente en los modelos utilizados en las estadísticas bayesianas. Discutimos la importancia de este resultado, así como un fenómeno que a menudo se desprende de las diferentes metodologías y aplicaciones prácticas de estos paradigmas, un fenómeno que llamamos efecto de Bayes.
Las estadísticas bayesianas difieren de las estadísticas frecuentistas en su tratamiento de valores desconocidos.
consideran la probabilidad como un concepto epistémico sin ninguna meta necesaria.
unida a supuestos de integridad de las preferencias y diversos supuestos de medición, implica la existencia de alguna creencia probabilística no degenerada para cualquier valor desconocido ajo este enfoque se dan parámetros desconocidos. Una distribución de probabilidad previa, que refleja las creencias previas y la incertidumbre del observador. Esta contrasta con el paradigma frecuentista donde los parámetros se consideran constantes desconocidas.

Se menciona que los modelos estándar bajo el paradigma bayesiano y frecuentista tratan los valores observables como infinitamente intercambiables, extensibles y por lo tanto independientes e idénticamente distribuidos (IID) condicionados en algún parámetro desconocido. Sin embargo, los diferentes tratamientos del parámetro desconocido conducen a diferentes resultados con respecto a la relación entre los valores observables. En particular, los frecuentistas consideran los valores observables como independientes, mientras que los bayesianos no lo hacen. En el artículo, se examina la relación entre intercambiabilidad y la correlación bajo cada paradigma y se discute un fenómeno denominado el efecto de Bayes.
Si x es intercambiable, entonces los elementos de x son independientes si y solo si la distribución empírica (o equivalente, cualquier otro parámetro que indexe la distribución empírica) es casi seguramente constante.
Los elementos de una secuencia intercambiable nunca pueden estar correlacionados negativamente y no están correlacionados solo cuando μ (θ) es casi constante.
Bajo el modelo frecuentista estándar, asumimos que los elementos de x son independientes con densidad de muestreo que depende del parámetro θ. Hemos demostrado que esto es equivalente a suponer que x es intercambiable junto con el tratamiento de θ que tiene una densidad de masa puntual en su propio valor verdadero θ T.

el frecuentista puede concluir que los elementos de x son independientes independientemente del valor verdadero de θ y, en particular, que x k + 1 y x k son independientes. Como tal, uno podría esperar que los valores observados no se usarían para predecir valores desconocidos. Sin embargo, dado que los frecuentistas toman el parámetro subyacente θ como una constante desconocida, no pueden obtener exactamente las densidades anteriores.
Este resultado puede parecer incongruente. Desde la independencia significa que el condicionamiento sobre lo observado los valores no afectan la medida de probabilidad para valores no observados, ¿cómo puede el frecuentista usar legítimamente el primero para predecir el último? La respuesta parece ser que el frecuentista acepta la independencia incondicional pero, sin embargo, está obligado a usar solo una estimación de la densidad predictiva.
Los bayesianos lo hacen en virtud de una afirmación explícita de dependencia (generalmente correlación positiva) entre estos valores. Los frecuentistas afirman que los valores son independientes pero que usan estos valores en la práctica para estimar las densidades desconocidas involucradas.

Aparte del hecho de que los bayesianos son capaces de hacer afirmaciones explícitas de correlación positiva entre las observaciones, hay otra diferencia que a menudo surge entre los paradigmas bayesiano y frecuentista. Al tratar con parámetros que creemos a priori probablemente están cerca de un valor particular, los bayesianos son capaces de acomodar esta creencia en la forma de una medida previa concentrada estrechamente en torno a este valor. Sin embargo, dado que los frecuentistas ven los parámetros como constantes, no pueden acomodar dicha información en ningún sentido probabilístico. En su lugar, se les deja una opción: descartar la información y tratar el parámetro como una constante desconocida como normal, o usar la información para tratar el parámetro como una constante conocida o una constante que se sabe que cae dentro de un rango estrecho.
El principio de predecir el resultado que se produce con mayor frecuencia en secuencias intercambiables no depende de la forma del modelo específico, procesos como el lanzamiento de monedas, que son diseñados para producir resultados

independientes y distribuidos uniformemente, pero que pueden estar sujetos a imperfecciones que conducen a sesgos; la ignorancia en cuanto a la dirección del sesgo debería llevarnos a Predecir el resultado que más se ha producido.

Los bayesianos son capaces de hacer afirmaciones explícitas de correlación positiva entre las observaciones y que existe otra diferencia que a menudo surge entre el enfoque bayesiano y frecuentistas. Este es que al tratar con parámetros que se creen prioritarios, probablemente se esta cerca a un valor particular.

4)
Dado que la sucesión X_t es dependiente markov entonces tenemos que $P(X_0=S_i, X_1=S_j, \ldots, X_T=S_t)$ se puede escribir como $P(X_0=S_i)P(X_1=S_j:X_0=S_i)*¿*P(X_T=S_t:X_{T-1}=S_k)$ para cadenas de markov no homogeneas se asume $X_{t-1}=S_i$ y que $X_t=S_j$ entonces $P(X_t=S_j:X_{t-1}=S_i)=P_{ij}(t)$ que no es otra cosa que la probabilidad de transición de pasar del estado t-1 al estado t, y N_{ij} es el número de veces tales que $X_{t-1}=S_i$ y $X_t=S_j$, este número de veces que se da la transición en el tiempo t se denota por la variables aleatorias $N_{ij}(t)$ y el vector aleatorio $N_i(t)=(N_{i1}(t), \ldots, N_{iJ}(t))$ el cual se distribuye Mult~ $(n_i(t), p_{ij}(t))$ entonces la pdf se puede definir como $f_i(N_i(t):n_i(t), p_{ij}(t))=P(N_i(t)=n_i(t), \ldots, N_J(t)=n_J(t))$ que a su vez se puede reescribir como $f_i(n_{i1}(t), \ldots, n_{iJ}(t):n_i(t), p_{ij}(t))=\frac{n_i(t)!}{n_{i1}!(t), \ldots, n_{iJ}(t)!} p_{i1}(t)^{n_{i1}(t)} \ldots p_{iJ}(t)^{n_{iJ}(t)}$ donde

$n_i(t)=\sum_{j=1}^{J} n_{ij}(t)$ para los renglones de la matriz de transición la distribución multinomial del renglón i es independiente del renglón j para todo j distinto de k.

Cuando la probabilidad conjunta $f_i(N_i(t):n_i(t), p_{ij}(t))$ es considerada como una función de los parámetros dado un vector $N_i(t)$, no es otra cosa que la función de verosimilitud queremos encontrar los parámetros para los cuales la probabilidad el vector $n_i(t)=(n_{i1}(t), \ldots, n_{iJ}(t))$ es muy alta. Y utilizamos este valor como un estimador para los parámetros.

Si usamos los estimadores de los parámetros para los cuales la función de verosimilitud es máxima entonces sea

$n_i(t)=\sum_{j=1}^{J} n_{ij}(t)$ y asumiendo que cada renglón es independiente, entonces la función de verosimilitud esta dada por:

$$l(\tilde{p}_i(t):\tilde{n}_i(t))=\prod_{i=1}^{I}\left[\frac{n_i(t)!}{\prod_{j=1}^{J} n_{ij}(t)}\right]\prod_{j=1}^{J} p_{ij}(t)^{n_{ij}(t)} aprox \prod_{i=1}^{I}\prod_{j=1}^{J} p_{ij}(t)^{n_{ij}(t)}$$

Sea $\tilde{p}=(p_{i1}, \ldots, p_{iJ})$ el vector de probabilidades del i-[esimo renglón, este se puede resumir como un prior de la pdf que podemos llamar $g_i(\tilde{p}_i)$ y sea $f_i(\tilde{N}_i:\tilde{p}_i)$ la función de densidad conjunta por tanto tenemos que :
$f_i(\tilde{N}_i:\tilde{p}_i)=l_i(\tilde{N}_i:\tilde{p}_i)g_i(\tilde{p}_i)=g_i(\tilde{p}_i:\tilde{N}_i)g_i(\tilde{N}_i)$ entonces podemos decir que la pdf posterior para el paramétro vector $\tilde{p}_i$ dado $\tilde{N}_i$ se puede ver como: $g_i(\tilde{p}_i:\tilde{N}_i) aprox\, g_i(\tilde{p}_i) l_i(\tilde{N}_i:\tilde{p}_i)$.

Si suponemos que la distribución a priori se distribuye Dirchlet y además suponemos independencia entre los renglones entonces la pdf posterior se puede ver como un producto de dirchlets con parámetros $n_{ij}(t)+a_{ij}(t-1)$ como:

$$g_i(\tilde{p}_i(t):\tilde{n}_i(t)) aprox \prod_{i=1}^{I}\left(\gamma\frac{\sum_{j=1}^{J} n_{ij}(t)+a_{ij}(t-1)}{\prod_{j=1}^{J}\gamma(n_{ij}(t)+a_{ij}(t-1))}\right)\prod_{j=1}^{J} p_{ij}(t)^{n_{ij}(t)+a_{ij}(t-1)-1}$$

donde gamma es la función gamma y $a_{ij}'s$ son parámetros positivos.

Ahora la distribución predictiva posterior para el renglón i está dada por:

$$P_i(\tilde{n}_i(t+1):\tilde{n}_i(t))=\int f_i(\tilde{n}_i(t+1):\tilde{p}_i(t),\tilde{n}_i(t))g_i(\tilde{p}_i(t):\tilde{n}_i(t))d\,\tilde{p}_i(t)$$

Entonces usando Dirchlet como la distribución a priori para la distribución multinomial con J categorías podemos observar que la distribución posterior es:

$$P_i(\tilde{n}_i(t+1):\tilde{n}_i(t))=\int_0^1[\prod_{i=1}^{I}[\frac{n_i(t+1)!}{\prod_{j=1}^{J}n_{ij}(t+1)}]\prod_{j=1}^{J}p_{ij}(t+1)^{n_{ij}(t+1)}]*[\prod_{i=1}^{I}(\gamma\frac{\sum_{j=1}^{J}n_{ij}(t)+a_{ij}(t-1)}{\prod_{j=1}^{J}\gamma(n_{ij}(t)+a_{ij}(t-1))})\prod_{j=1}^{J}p_{ij}(t)^{n_{ij}(t)+a_{ij}(t-1)-1}]d$$

$$¿\frac{n_i(t+1)!}{\prod_{j=1}^{J}n_{ij}(t+1)!}*(\gamma\frac{n_{ij}(t)+a_{ij}(t-1)}{\prod_{j=1}^{J}\gamma(n_{ij}(t)+a_{ij}(t-1))})*(\frac{\prod_{j=1}^{J}\gamma(n_{ij}(t+1)+n_{ij}(t)+a_{ij}(t-1))}{\gamma n_{ij}(t)+a_{ij}(t-1)})$$

donde $n_i(t+1)=\sum_{j=1}^{J}n_{ij}(t+1),\quad n_i(t)=\sum_{j=1}^{J}n_{ij}(t)$ **y** $a_i(t-1)=\sum_{j=1}^{J}a_{ij}(t-1)$