Tr. J. of Veterinary and Animal Sicences
22 (1998) 485–488
© TÜBİTAK

Hayvan Islahı Uygulamalarında Kullanılan İnvers-χ^2 ve İnvers-Wishart Dağılışlarının Matematik Esasları

M. Ziya FIRAT, Yüksel BEK
Çukurova Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Adana-TÜRKİYE

Geliş Tarihi: 20.09.1996

Özet: Ki-kare ve Wishart dağılışları hayvan ıslahı uygumalarında özellikle Bayesian metodu kullanılarak varyans unsurlarını tahmin etmede önemli bir oynarlar. Bu iki dağılış, biri diğerinden kolayca türetilebilecek kadar ilişkilidirler. Diğer bir ifade ile, ki-kare dağılışı Wishart dağılışının özel bir durumudur veya Wishart dağılışı ki-kare dağılışının çok değişkenli hale genelleştirilmesidir. Bu makale yukarıda bahsedilen dağılışların basit bir matematik esasını verir, birinin diğerinden nasıl türetildiğini ve hayvan ıslahı uygulamalarında ne şekilde kullanıldıklarını gösterir.

Anahtar Sözcükler: Ki-kare dağılışı, Wishart dağılışı, hayvan ıslahı, varyans unsurları

The Mathematical Basis of Inverseχ^2 and Inverse-Wishart Distribution Used in Animal Breeding Applications

Abstract: The chi-square and Wishart distributions play an important role in animal breeding applications, especially in the estimation of variance components using Bayesian methods. These two distributions are so related that one can easily be derived from the other. In other words, the chi-square distribution is a particular case of the Wishart distribution or the Wishart distribution is a multivariate generalization of the chi-square distribution. This article provides a simple mathematical basis of the above-mentioned distributions; it shows how one is derived from the other and in what from they are used in animal breeding applications.

Key Words: Chi-square distribution, Wishart distribution, animal breeding, variance components

Giriş

Wishart dağılışı ile χ^2 dağılışı arasında sıkı bir ilişki olup, Wishard dağılışı χ^2 dağılışının çoklu karakterler için genelleştirilmiş halidir. Bu dağılışların konjugate olma özellikleri nedeniyle yapılacak matematiksel analizler önemli ölçüde basitleştiğinden Bayesian metodları ile varyans unsurlarının tahmininde prior dağılışı olarak sıkça kullanılmaktadır. Zellner (1), Fouller ve ark. (2) ve Cantet ve ark. (3) varyans ve kovaryans matrislerinin prior dağılışı için invers Wishart dağılışı kullanmışlardır. Benzer şekilde Cantet ve ark. (3) ve Hoeshele ve ark. (4) hata varyans unsuru için invers χ^2 dağılışını önermişlerdir. Tek yönlü bir tesadüf modeli kullanılarak yapılan bir çalışmada varyans unsurlarının her ikisinin prior dağılışının bağımsız invers χ^2 dağılışına sahip oldukları varsayılmıştır (5, 6, 7). Herhangi bir varyans unsuru için invers χ^2 dağılışının seçimi bu dağılışın konjugate yapısından kaynaklanmaktadır (1, 6, 8).

Son yıllarda özellikle Bayesian metodunun hayvan ıslahında kullanılmaya başlamasından sonra invers χ^2 ve invers Wishart dağılışları oldukça önem kazanmaya başlamıştır. Hayvan ıslahı uygulamalarında sıkça kullanılan boğa ve hataya ait varyans unsurları, σ_s^2 ve σ_e^2 , veya bunların çok değişkenli halleri olan $\sum_s$ ve $\sum_e$ matrisleri ile tanımlanmaktadır. Bayesian analizinde bu parametreler için belirlenen uygun prior (ön) dağılışları genellikle invers χ^2 veya invers Wisharttır. Bu makalenin amacı, genellikle χ^2 ve Wishart dağılışlarının matematiksel esaslarını vermek, birbirlerinden nasıl elde edildiklerini ve aralarındaki ilişkileri göstermek ve hayvan ıslahı uygulamalarında bu dağılışların kullanımını incelemektir.

Ki-kare ve İnverst Ki-kare Dağılışları

Eğer X değişkeni aşağıda verilen olasılık yoğunluk fonksiyonuna sahip bir tesadüf değişkeni ise

$$f(\chi|v)=\frac{(\chi)^{\frac{v}{2}-1}\exp\left(-\frac{1}{2}\chi\right)}{\Gamma\left(\frac{v}{2}\right)2^{\frac{v}{2}}}, \quad \chi>0. \qquad (1)$$

bu durumda X'in, v serbestlik dereceli bir χ^2 dağılışına sahip olduğu söylenir ve $X\sim\chi^2_v$ ile gösterilir. X'in ortalama ve varyansının $E(X) = v$ ve $Var(X) = 2v$ olduğu kolaylıkla gösterilebilir. Bayesian analizinde X'in tersi olan X^{-1} doğal olarak ortaya çıkmaktadır. v serbestlik derecesine sahip *invers* χ^2 dağılışı, $Y = X^{-1} = 1/X$ transformasyonu

yapılarak (1) nolu eşitlikten türetilir ve dağılış fonksiyonu aşağıdaki şekilde tanımlanmaktadır.

$$f(y|v)=\frac{(y)^{\frac{v}{2}+1}\exp\left(-\frac{1}{2}y\right)}{\Gamma\left(\frac{v}{2}\right)2^{\frac{v}{2}}}, \qquad y>0. \qquad (2)$$

Bu dağılışın ortalama ve varyansı sırasıyla E(Y) = 1/(v-2) ve Var(Y) = 2/(v-2)^2(v-4) olarak verilebilir ve $X\sim\chi_v^{-2}$ ile gösterilir.

Eğer Y = SX olsaydı, bu durumda $1/Y\sim S^{-1}\chi_v^2$ şeklinde dağılış gösterirdi. Diğer bir ifade ile, Y, $Y\sim S\chi_v^{-2}$ ile gösterilen v serbestlik dereceli invers Ki-kare dağılışının S katı bir dağılışa sahip olacağından yoğunluk fonksiyonuda aşağıdaki gibi tanımlanırdı

$$f(y|v)=\frac{(y)^{\frac{v}{2}-1}S^{\frac{v}{2}}\exp\left(-\frac{S}{2y}\right)}{\Gamma\left(\frac{v}{2}\right)2^{\frac{v}{2}}}, \qquad y>0. \qquad (3)$$

Ortalama ve varyans sırasıyla E(Y) = S/(v-2) ve Var(Y)=2S^2/(v-2)^2(v-4) şeklinde verilir.

Wishart ve İnvers-Wishart Dağılışları

X matrisi, t(t+1)/2 adet farklı tesadüf değişkeninden, x_{ij}(ij=1,...t;i≥j) ibaret t x t boyutlu pozitif tanımlı simetrik bir tesadüf matrisi olsun. v≥t ve Σt x t pozitif tanımlı simetrik bir matrisi olsun. x_{ij}'nin dağılışı χ^2 dağılışının çok değişikli durumu için genelleştirilmiş halidir ve aşağıdaki gibi verilebilir

$$f(X|v,\ )=c^{-1}|X|^{(v-t-1)/2}\exp\left[-\frac{1}{2}tr(\ ^{-1}X)\right],|X|>0, \qquad (4)$$

burada $c^{-1}=2^{vt/2}\pi^{t(t-1)/4}|\Sigma|^{v/2}\Gamma_p\left(\frac{1}{2}v\right)$ ve $\Gamma_p\left(\frac{1}{2}v\right)=\pi^{t(t-1)/4}\prod_{i=1}^{t}\Gamma\left[\frac{1}{2}(v+1-i)\right]$

$\Gamma_p(.)$ çok değişkenli gamma fonksiyonu olarak adlandırılır. (4) nolu eşitlikteki dağılış $W_t(\Sigma,v)$ ile gösterilir ve X, v serbestlik dereceli ve parametresi Σ matrisi olan Wishart dağılışına sahiptir denir.

Eğer t=1 ve Σ=1 ise, Wishart dağılışı (1) nolu eşitlikte verilen v serbestlik dereceli χ^2 dağılışı haline gelir.

İnverse Wishart dağılışı X'in inversi alındıktan sonra Jacobian kullanılarak elde edilebilir. x^{ij} X'in inversinin (i, j)'inci elemanını göstersin. Bu durumda t(t+1)/2 tesadüf değişkeninin (x_{11}, x_{12},...,x_{tt})'den (x^{11}, x^{12},...,x^{tt})'ye transformasyonunun Jacobian'ı aşağıdaki gibidir.

$$J=\left|\frac{\partial(\chi_{11},\chi_{12},\ldots\chi_{tt}}{\partial(\chi^{11},\chi^{12},\ldots\chi^{tt}}\right|=|X|^{t+1}$$

Sonuç olara, inverst Wishart dağılışının olasılık yoğunluk fonksiyonu şöyle

$$f(X^{-1}|v,\Sigma)=c^{-1}|X|^{(v+t+1)/2}\exp\left[-\frac{1}{2}tr(\chi^{-1}\Sigma)\right], \quad |X|>0, \quad (5)$$

olur. Böylece (4) nolu eşitlikteki X'in dağılışı, v serbestlik dereceli t-boyutlu inverst Wishart dağılışı olarak adlandırılabilir ve $X^{-1}\sim W\ (\Sigma,v)$ ile gösterilir. Burada Σ ve v'ye *hiperparametreler* adı verilir.

t=1 olduğu zaman (5) nolu eşitlikteki dağılış (2)'deki *invers* χ^2 dağılışına indirgenir. (5) nolu dağılıştaki X matrisinin x_{il} gibi tek bir özelliği ile ilgili olasılık yoğunluk fonksiyonu şöyle olur

$$f(\chi_{i1}|v,\sigma_{11})\propto(\chi_{i1})^{-\frac{1}{2}(v+2)}\exp\left(-\frac{1}{2}\frac{\sigma_{11}}{\chi_{i1}}\right), \quad \chi_{i1}>0.$$

Ki-kare ve Wishart Dağılışlarının Hayvan Islahı Uygulamalarında Kullanımı

Son beş yıldır parametre tahmini ve damızlık değeri hesaplamaları için hayvan ıslahında Bayesian metodları giderek yaygınlaşmaktadır. Bayesian analizi, veriler için parametrik bir model seçip, bilinmeyen parametrelere prior (ön) dağılış belirleyip daha sonra bu modelin olabilirlik fonksiyonu ile prior dağılışların birleşiminden meydana gelen müşterek posterior (son) dağılışı incelemekten ibarettir. Parametrelerin prior dağılışı hayvan ıslahçısının parametre değerleri hakkındaki ön bilgisini doğru bir şekilde yansıtmalıdır. Ön bilgiyi analize dahil eden istatistiksel yöntemler bu ön bilgiyi dikkate almayan yöntemlere oranla daima daha kesin ve doğru yorumlamalar yapılmasını sağlarlar. Bu durum Bayesian yönteminin hayvan ıslahındaki önemini ortaya koymaktadır. Bu bölümde hayvan ıslahında Bayesian metodu uygulanırken invers-χ^2 ve invers-Wishart dağılışlarının nasıl ve hangi formda kullanıldıkları tartışılacaktır.

Önce tek yönlü boğa modelini ele alalım. Bu modelde-

ki varyans unsurları boğa ve hata varyansları, σ_s^2 ve σ_e^2, olup bu parametreler için prior dağılışın formüle edilmesi gerekmektedir. Bunu incelemeden önce boğalar arası ve boğalar içi kareler toplamlarının, S_b ve S_w, dağılışlarını incelemede yarar vardır. Boğalar arası ve boğalar içi kareler toplamları aşağıdaki gibi olup

$$S_b = n\sum_{i=1}^{s} (\bar{y}_{i.} - \bar{y}_{..})^2 \quad \text{ve} \quad S_w = \sum_{i=1}^{s}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2,$$

sırasıyla $(\sigma_e^2 + n\sigma_s^2)\chi^2_{s-1}$ ve $\sigma_e^2\chi^2_{s(n-1)}$ şeklinde χ^2 dağılışı gösterirler, yani,

$$f(S_b/\sigma_s^2, \sigma_e^2) = \frac{(\sigma_e^2 + n\sigma_s^2)^{-\frac{1}{2}(s-1)} (S_b)^{\frac{1}{2}(s-1)-1} \exp\left(-\frac{1}{2}\frac{S_b}{(\sigma_e^2 + n\sigma_s^2)}\right)}{\Gamma\left(\frac{1}{2}(s-1)\right) 2^{\frac{1}{2}(s-1)}}$$

ve

$$f(S_w/\sigma_e^2) = \frac{(\sigma_e^2)^{-\frac{1}{2}s(n-1)} (S_w)^{\frac{1}{2}s(n-1)-1} \exp\left(-\frac{1}{2}\frac{S_w}{2\sigma_e^2}\right)}{\Gamma\left(\frac{1}{2}s(n-1)\right) 2^{\frac{1}{2}s(n-1)}}$$

Varyans unsurları, σ_s^2 ve σ_e^2 'nin değerleri hakkındaki bilgiyi temsil eden prior dağılışı formüle etmek için değişik girişimlerde bulunulmuştur. Burada önemli olan subjektif inancı matematiksel fonksiyon şeklinde ifade etmektedir. Hayvan ıslahı uygulamalarında bu parametreler için en çok kullanılan prior dağılış, invers χ^2 olup sırasıyla aşağıdaki gibi verilebilir

$$f(\sigma_s^2 | \nu_s, s_s^2) \propto (\sigma_s^2)^{-\frac{1}{2}(\nu_s+2)} \exp\left(-\frac{V_s S_s^2}{2\sigma_s^2}\right), \quad 0 \le \sigma_s^2 \le \sigma_e^2/3,$$

ve

$$f(\sigma_e^2 | \nu_e, s_e^2) \propto (\sigma_e^2)^{\frac{1}{2}(\nu_e+2)} \exp\left(-\frac{V_e S_e^2}{2\sigma_e^2}\right), \quad 0 < \sigma_e^2 < \infty,$$

burada s_s^2 ve s_e^2, σ_s^2 ve σ_e^2 'nin prior beklenen değerleri olarak adlandırılabilir ve ν_s ve ν_e varyans unsurlarının pior değeri hakkındaki inancın derecesini yansıtan serbestlik derecesi niteliğinde tamsayı değerinde parametrelerdir. Bu dört parametreye *hiperparametreler* adı verilir.

Bundan sonraki kısımda buraya kadar anlatılanlar çok değişkenliler için genelleştirilmektedir. Tek değişkenli analizden elde edilen boğalar arası ve boğalar içi kareler toplamları S_b ve S_w, çok değişkenli analizde boğalar arası ve boğalar içi kareler ve çarpımlar toplamı matrisleri, S_b ve S_w, olarak isimlendirilerek aşağıdaki gibi verilebilir

$$S_b = n\sum_{i=1}^{s} (\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{i.} - \bar{y}_{..})' \text{ ve } S_w = \sum_{i=1}^{s}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})(y_{ij} - \bar{y}_{i.})'$$

Yukarıda verilen kareler ve çarpımlar toplamı matrisleri birbirinden bağımsız olup sırasıyla $S_b \sim W_t[s-1, \Gamma_n]$ ve $S_w \sim W_t[s(n-1), \Sigma_e]$ şeklindeki merkezi Wishart dağılışı gösterirler. Buna göre S_b ve S_w'nin yoğunluk fonksiyonları aşağıdaki gibi yazılabilir

$$f(S_b | \Gamma_n) \propto |\Gamma_n|^{-\frac{1}{2}(s-1)} \exp\left[-\frac{1}{2} tr(S_b \Gamma_n^{-1})\right]$$

ve

$$f(S_w | \Sigma_e) \propto |\Sigma_e|^{-\frac{1}{2}s(n-1)} \exp\left[-\frac{1}{2} tr(S_w \Sigma_e^{-1})\right],$$

burada $\Gamma_n = \Sigma_e + n\Sigma_s$, Σ_s boğa varyans-kovaryans matrisi ve Σ_e hata varyans-kovaryans matrisidir.

Bayesian yönteminde Σ_s ve Σ_e matrisleri hakkındaki mevcut prior bilgilerin analize dahil edilmesi gerekir. Örneğin bu prior bilgi aynı ırktaki hayvanlardan daha önce elde edilen verilerden sağlanabilir. Varyans kovaryans matrisleri için en uygun konjugate prior dağılış invers Wishart'tır. Bundan dolayı Σ_s ve Σ_e'nin prior dağılışlarını invers Wishart dağılışı olarak varsaymak uygun olabilir (1, 2) ve bunlar aşağıdaki gibi verilmiştir

$$f(\Sigma_s | \nu_s, S_s) \propto |\Sigma_s|^{-\frac{1}{2}(\nu_s+t+1)} \exp\left[-\frac{1}{2} tr(\Sigma_s^{-1} S_s)\right],$$

ve

$$f(\Sigma_e | \nu_e, S_e) \propto |\Sigma_e|^{-\frac{1}{2}(\nu_e+t+1)} \exp\left[-\frac{1}{2} tr(\Sigma_e^{-1} S_e)\right],$$

burada t karakter (değişken) sayısı, S_s ve S_e, $\sum_s$ ve $\sum_e$'nin prior dağılışlarının beklenen değerlerine ait txt boyutlu matrisler ve v_s ve v_e ise S_s ve S_e'ye olan inanç derecesinin bir ölçüsü veya serbestlik derecesi olarak isimlendirilen tamsayı değerlerdir.

Bu makale ki-kare ve Wishart dağılışlarının özellikleri incelenerek, hayvan ıslahı uygulamalarında bunlardan nasıl yararlanıldıkları gösterilmiştir. Burada bilinmesi gereken en önemli husus, Wishart'ın ki-kare dağılışının özel bir hale olup, değişken sayısı bire eşit olduğu zaman ki-kareye indirgenmesidir. Bu dağılışların ve terslerinin hayvan ıslahı uygulamalarında kullanımı oldukça önem arzetmektedir. Özellikle tek değişkenli modellerdeki varyans unsurlarının ve çok değişkenli modellerdeki varyans kovaryans matrislerinin prior dağılışlarının ki-kare ve Wishart dağılışları kullanılarak nasıl formüle edildikleri üzerinde durulmuştur. Asıl incelenmesi gereken bir başka konu da, Bayesian analizinde invers χ^2 ve invers Wishart dağılışlarından nasıl tesadüf sayıları elde edildiğidir. Ki-kare için tesadüf sayıları üreten programlar mevcut olmasına ragmen, Wishart dağılışı bilgisayar simülasyon algoritmaları yaygın olarak kullanıma açık olmayan birkaç çok değişkenli dağılıştan bir tanesidir. Bundan dolayı Wishart dağılışının simülasyonunun nasıl gerçekleştirildiği biyometri ve genetik-ıslah çalışmalarında önemil olup başka bir araştırmada incelenmesi uygun olacaktır.

Kaynaklar

1. Zellner, A., An Introduction to Bayesian Inference in Econometrics. Wiley, N.Y., 1971.
2. Foulley, J.L., Im, S., Gianola, D. and Hoeschele, I., Empirical Bayes estimation of parameters for n polygenic binary traits. Genet. Sel. Evol. 1987, 19: 197-224.
3. Cantet, R.J.C., Fernando, R.L. and Gianola, D., Bayesian inference about dispersion parameters of univariate mixed models with maternal effects: Theoretical considerations. Genet. Sel. Evol. 1992, 24: 107-135.
4. Hoeschele, I., Gianola, D., and Foulley, J.L., Estimation of variance components with Quassi-continuous data using Bayesian methods. J. Anim. Breed. and Genet. 1987, 104: 334-349.
5. Lindley, D.V. and Smith, A.F.M., Bayes estimation for the linear model. J. Roy. Statists. Soc. Ser. B. 1972, 34: 1-18.
6. Broemeling, L.D., Bayesian Analysis of Linear Models. Marcel Dekker, N.Y., 1985.
7. Gianola D., Im, S. and Macedo, F.W., A framework for prediction of breeding value. In: Advances in Statistical Methods for the Genetic Improvement of Livestock (Gianola, D. and Hammond, K. eds) Springer-Verlag, Berlin. 1990, 210-238.
8. Box, G.E.P. and Tiao, G.C., Bayesian Inference in Statistical Analysis. Addison-Weiley, Reading, 1973.